

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 18:17:26 ; Search time 400 Seconds  
(without alignments)  
6251.504 Million cell updates/sec

Title: US-09-386-850-7\_COPY\_1\_510

Perfect score: 510  
Sequence: 1 AACTGACTGACGAGAGC.....TTTTCGAGGTGTGGCAA 510

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3222919 seqs, 2451570024 residues

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq2:  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	510	100.0	658	10	US-09-386-850-7
2	65	12.7	65	10	US-09-908-975-3241
3	53.4	10.5	153	9	US-09-919-580-624
4	39.4	7.7	65	10	US-09-908-975-29028
5	39.4	7.7	56153	13	US-10-221-714A-519
6	38.2	7.5	1391	17	US-10-713-836-5
7	38.2	7.5	6022	13	US-10-221-714A-383
8	38.2	7.5	40304	12	US-09-997-722-280
9	37.4	7.3	19653	15	US-10-311-455-1308
10	36.8	7.2	630	17	US-10-437-963-78930
11	36.6	7.2	3673778	15	US-10-312-841-2
12	36.4	7.1	5127	15	US-10-239-676-132
13	36.4	7.1	6222	15	US-10-311-455-666
14	36.2	7.1	777	13	US-10-027-632-135035

15	36.2	7.1	777	13	US-10-027-632-135036	Sequence 135036,
16	36.2	7.1	777	13	US-10-027-632-135037	Sequence 135037,
17	36.2	7.1	777	16	US-10-027-632-135035	Sequence 135035,
18	36.2	7.1	777	16	US-10-027-632-135036	Sequence 135036,
19	36.2	7.1	777	16	US-10-027-632-135037	Sequence 135037,
20	36.2	7.1	5815	17	US-10-257-166-105	Sequence 105, App
21	36.2	7.1	8776	17	US-10-257-166-150	Sequence 150, App
22	36.2	7.1	659158	9	US-09-771-208-20	Sequence 20, Appl
23	36	7.1	644	9	US-09-759-143-752	Sequence 752, App
24	36	7.1	644	9	US-09-780-669-752	Sequence 752, App
25	36	7.1	644	9	US-09-822-827-752	Sequence 752, App
26	36	7.1	644	9	US-09-895-793-752	Sequence 752, App
27	36	7.1	644	9	US-09-895-814-752	Sequence 752, App
28	36	7.1	644	14	US-10-012-896-752	Sequence 752, App
29	36	7.1	644	15	US-10-144-678A-752	Sequence 752, App
30	36	7.1	644	15	US-10-294-025-752	Sequence 752, App
31	35.8	7.0	7029	15	US-10-311-455-593	Sequence 593, App
32	35.6	7.0	549	13	US-10-621-901-266	Sequence 266, App
33	35.6	7.0	4255	15	US-10-311-455-2006	Sequence 2006, Ap
34	35.6	7.0	3673778	15	US-10-312-841-1	Sequence 1, Appli
35	35	6.9	14253	15	US-10-311-455-1467	Sequence 1467, Ap
36	34.8	6.8	24259	13	US-10-221-714A-415	Sequence 415, App
37	34.6	6.8	65	10	US-09-908-975-27963	Sequence 27963, A
38	34.4	6.7	523	9	US-09-864-761-30606	Sequence 30606, A
39	34.4	6.7	1721	13	US-10-027-632-98304	Sequence 98304, A
40	34.4	6.7	1721	16	US-10-027-632-98304	Sequence 98304, A
41	34.4	6.7	10039	15	US-10-311-455-2015	Sequence 2015, Ap
42	34.4	6.7	33578	16	US-10-085-117-238	Sequence 238, App
43	34.2	6.7	1150	13	US-10-115-123-69	Sequence 69, Appl
44	34.2	6.7	1150	15	US-10-012-542-69	Sequence 69, Appl
45	34.2	6.7	1310	17	US-10-437-963-60606	Sequence 60606, A

## ALIGNMENTS

RESULT 1  
US-09-386-850-7

; Sequence 7, Application US/09386850  
; Publication No. US20030170620A1

; GENERAL INFORMATION:

; APPLICANT: Rosinski-Chupin, Isabelle R.

; Ironik, Diana

; Rougeon, Francois

; Seidah, Nabil

; TITLE OF INVENTION: Peptides and Polypeptides Derived

; from the Submaxillary Gland of the Rat, Corresponding  
; Antibodies, Corresponding Hybridomas, and Uses of  
; These Products for Diagnosis, Detection, or

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Burns, Doane, Swecker & Mathis

; STREET: P.O. Box 1404

; CITY: Alexandria

; STATE: VA

; COUNTRY: USA

; ZIP: 22313-1404

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/386,850

; FILING DATE: 31-Aug-1999

; CLASSIFICATION DATA:

; APPLICATION NUMBER: US/08/476,120

; FILING DATE: 07-JUN-1995

; APPLICATION NUMBER: US 08/153,277

; FILING DATE: 17-NOV-1993

; APPLICATION NUMBER: US 07/499,276

; FILING DATE: 19-JUL-1990

```
; APPLICATION NUMBER: PCT/FR88/00523
; FILING DATE: 11-OCT-1989
; APPLICATION NUMBER: FR 88/13353
; FILING DATE: 11-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 004900-129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-8620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 658 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: No. US20030170620A1 Relevant
; TOPOLOGY: No. US20030170620A1 Relevant
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Rat
; TISSUE TYPE: Submaxillary Gland
; CELL TYPE: Glandular
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: SMR1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 73..510
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-386-850-7

Query Match      100.0%; Score 510; DB 10; Length 658;
Best Local Similarity 100.0%; Pred. No. 1.1e-151;
Matches 510; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAACAGCTGACGACAGAGAGCTTCTGACGACGACATTTCCCGCTCAGAGATTCTCCCAAG 60
Db 1 AAACAGCTGACGACAGAGAGCTTCTGACGACGACATTTCCCGCTCAGAGATTCTCCCAAG 60

QY 61 GGGCTACCAAGATGAAGTACATGATTTGATCTTTGGCTGTGGATCTCTAGCATGC 120
Db 61 GGGCTACCAAGATGAAGTACATGATTTGATCTTTGGCTGTGGATCTCTAGCATGC 120

QY 121 TTCCAGTTCAGGTGAGGTGTGACAGGCCCAAGAGACAACATAATCTTAGAGACAACAA 180
Db 121 TTCCAGTTCAGGTGAGGTGTGACAGGCCCAAGAGACAACATAATCTTAGAGACAACAA 180

QY 181 GATCCTTCAACTCTTCTCATTAATCTTGCTTCAGCCTGATCCCAATGGTGACAAATA 240
Db 181 GATCCTTCAACTCTTCTCATTAATCTTGCTTCAGCCTGATCCCAATGGTGACAAATA 240

QY 241 GGAGTAACAATCACTATACCTTAAATCTTCAACCACTGCTGTCTTGTGTAATCTTCCC 300
Db 241 GGAGTAACAATCACTATACCTTAAATCTTCAACCACTGCTGTCTTGTGTAATCTTCCC 300

QY 301 GGTGTTTATCACTGGACACCAATGGTTGTACAAGGTACCACTGAATATCAATATCATGTGG 360
Db 301 GGTGTTTATCACTGGACACCAATGGTTGTACAAGGTACCACTGAATATCAATATCATGTGG 360

QY 361 CAGCTAACTGCTCCAGACCTACACCTCTAGCAATCTTCTACTCAATCTTATCTCCACA 420
Db 361 CAGCTAACTGCTCCAGACCTACACCTCTAGCAATCTTCTACTCAATCTTATCTCCACA 420

QY 421 GAACAAGCAAAATACAAAAACAGATGCCAAAATCTCCAACACTACTGCGACTACCCAAAAT 480
Db 421 GAACAAGCAAAATACAAAAACAGATGCCAAAATCTCCAACACTACTGCGACTACCCAAAAT 480

QY 481 TCCACTGATATTTTGAAGTGTGGCAAA 510
Db 481 TCCACTGATATTTTGAAGTGTGGCAAA 510

RESULT 2
US-09-908-975-3241
```

```
; Sequence 3241, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASSERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Liat
; APPLICANT: FAIGLER, Simchon
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
; TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME
; FILE REFERENCE: 36888-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 3241
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; US-09-908-975-3241

Query Match      12.7%; Score 65; DB 10; Length 65;
Best Local Similarity 100.0%; Pred. No. 2.1e-10;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 277 CTCGTGTTCTTGTAAATCTTCCCGTTTATCACTGGACCACTTGTGTACAAGT 336
Db 1 CTCGTGTTCTTGTAAATCTTCCCGTTTATCACTGGACCACTTGTGTACAAGT 60

QY 337 ACCAC 341
Db 61 ACCAC 65

RESULT 3
US-09-919-580-624
; Sequence 624, Application US/09919580
; Patent No. US20020110832A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.552
; CURRENT APPLICATION NUMBER: US/09/919,580
; CURRENT FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 624
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-919-580-624

Query Match      10.5%; Score 53.4; DB 9; Length 153;
Best Local Similarity 63.8%; Pred. No. 1.8e-06;
Matches 81; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 69 AAAGATGAAGTCACTGATTTTGTATCTTTGGCCTGTGGATCTTCTAGCATGCTTCAGTC 128
Db 26 AAAGATGAAGTCACTGATTTTGTATCTTTGGCCTTTGGCCTTTTGCAGCGTGTTCACACC 85

QY 129 AGGTAGGGTGTGAGAGGCCCAAGAGACAACATATCTTAGAAGACAACAGATCTTC 188
Db 86 TGGTGAAGTCAAGAGAGGCCCAAGAGACCATATCCACTGGACCGCTGCTCTCTCTCA 145

RESULT 2
US-09-908-975-3241
```

```
Db      146 ACCTTTT 152

RESULT 4
US-09-908-975-29028
; Sequence 29028, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASSERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Liat
; APPLICANT: FAIGLER, Simchon
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
; TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME
; FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 29028
; LENGTH: 65
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-908-975-29028

Query Match      7.7%; Score 39.4; DB 10; Length 65;
Best Local Similarity 75.4%; Pred. No. 0.031;
Matches 49; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY      424 CAAGCAATAACAAAACAGATGCAAAATCTCCAAACACTACTGCGACTACCCAAAATTCC 483
Db      1 CAAGCAACTACAATGCCAGTGCAGTATCTCCATTACTACTCTCTACTGCTAGAGATTCC 60

QY      484 ACTGA 488
Db      61 ACTGA 65

RESULT 5
US-10-221-714A-519/c
; Sequence 519, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 519
; LENGTH: 56153
; TYPE: DNA
; ORGANISM: Artificial Sequence
```

```
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-519

Query Match      7.7%; Score 39.4; DB 13; Length 56153;
Best Local Similarity 50.8%; Pred. No. 1.9;
Matches 94; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY      149 CAAGAAGACAAACATAATCTCTAGAAGACAAACAGATCTTCAACTCTTCCTCAATTATCTTG 208
Db      11877 CAACATCTAACAAAATATACACGAAAAACACCTTAAAAAACCCTTAACACATAT 11818

QY      209 GTCTTCAGCCTGATCCCAATGGTGGCAATAGAGATTAACAATCACTATACCTTAAATC 268
Db      11817 AACACTAAACTAATAATAATAAAAAAACAACATCTCGAATTCACAAAAAC 11758

QY      269 TTCAACCCACCTCGTGTCTTGTAACTCTCCCGGTTTATCACTGGACCACTTGGTTG 328
Db      11757 ATCCCTCTCTCGAAAACTATTCAATCCCTATATTTTAACTCTTCTCTACTAATCATCG 11698

QY      329 TACAA 333
Db      11697 TACCA 11693

RESULT 6
US-10-713-836-5/c
; Sequence 5, Application US/10713836
; Publication No. US20040143871A1
; GENERAL INFORMATION:
; APPLICANT: Dhugga, Kanwarpal S.
; TITLE OF INVENTION: Isolation of Genes for Galactomannan
; TITLE OF INVENTION: Formation from Guar Seeds
; FILE REFERENCE: 1520
; CURRENT APPLICATION NUMBER: US/10/713,836
; CURRENT FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/426,127
; PRIOR FILING DATE: 2002-11-14
; PRIOR APPLICATION NUMBER: 60/490,022
; PRIOR FILING DATE: 2003-07-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1391
; TYPE: DNA
; ORGANISM: Cyamopsis tetragonoloba
US-10-713-836-5

Query Match      7.5%; Score 38.2; DB 17; Length 1391;
Best Local Similarity 58.3%; Pred. No. 0.48;
Matches 67; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY      189 AACTCTCTCATTAATCTTGGTCTTCAGCTGATCCCAATGGTGGACAAATAGGAGTAAC 248
Db      1018 ACCTCTCGTTATGTTCTCTGAGCTTTCACGATCTCCAGCAGTACCCCTTCAAGTAGT 959

QY      249 AATCACTATACCCCTTAAATCTTCAACCACTGCTGTCTTGTGTTTAAATCTTCCCGGT 303
Db      958 ACTCGCTCTCCAAGAAGATCTTCTCCCTCCAGCTGCTTATATATCCGTCGGAT 904

RESULT 7
US-10-221-714A-383/c
; Sequence 383, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
```



```
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 78930
; LENGTH: 630
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_7868C.1
US-10-437-963-78930

Query Match          7.2%; Score 36.8; DB 17; Length 630;
Best Local Similarity 48.6%; Pred. No. 0.83;
Matches 101; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 157 CRACATAATCTAGAGACAGACAGATCTTCACTCTCTCTCATATCTGTGCTTTCAG 216
Db 137 CATCATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 196

QY 217 CTGATCCCAATGGTGACAAATAGGAGTAACAATCACTATACCTTAAATCTTCAACCA 276
Db 197 CGTCATCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 256

QY 277 CTCCTGTTCTTGTAAATCTTCCCGTTTATACATGGACCACTTGGTTGTACAAAGT 336
Db 257 CTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 316

QY 337 ACCACTGAATATCAATATCAGTGGCAGC 364
Db 317 CACTCTCCATTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 344

RESULT 11
US-10-312-841-2/c
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (379615)
US-10-312-841-2

Query Match          7.2%; Score 36.6; DB 15; Length 3673778;
Best Local Similarity 50.3%; Pred. No. 1.7e+02;
Matches 90; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 303 TTTTATCACTGGCCACCATTGGTTGTACAGTACCACTGAATATCAATATCAGTGGCA 362
Db 308899 TTTTATTTTCAAAAACCGCAATTTATAACTTAACCAATTTTACCAAACTAACTTAAACA 308840

QY 363 GCTAATCTCTCCAGACCTTACACCTCTAAGCAATCTCTCTCTCTCTCTCTCTCTCTCT 422
Db 308839 ACNAATCTTAAATCTTAAACCCACATTTTATCTTATTAACCCCTTTTAAACCAA 308780

QY 423 ACAAGCAATATACAAAACAGATGCCAAAATCTCCAACTACTCGCACTACCCAAAATT 481
Db 308779 TAACACAAAACCAATTCATATACACAAATATACACCAATTTTACTACAACTTAAAT 308721

RESULT 12
US-10-239-676-132/c
; Sequence 132, Application US/10239676
; Publication No. US20030082609A1
```

```
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 132
; LENGTH: 5127
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (1976, 1981...1982, 1984, 1989...1990, 1995...1996, 5032, 5059)
US-10-239-676-132

Query Match          7.1%; Score 36.4; DB 15; Length 5127;
Best Local Similarity 52.0%; Pred. No. 4;
Matches 106; Conservative 0; Mismatches 96; Indels 2; Gaps 1;

QY 257 TACCCCTTAAATCTTCAACCACTCGTGTCTTCTTAACTCTCCGGGTTTATCACTGGAC 316
Db 2577 TACGATTTTCTCTATCATTTCTCTCTCATTTTAAATCCCAATCTCTCTCAACA 2518

QY 317 CACCATTTGGTTGTACAAAGGTACCACTGAATATCAATATCATGTGGCAGCTAATCTCTCCAG 376
Db 2517 AAACAATAACAAAAAACCTTAACTTTATAAAAAAACAACACCCCTCTCTCCAA 2458

QY 377 ACCTACACCTCTAAGCAATCTCTCTCTCACTCACTTCACTCCAGAACAGCAATATCAA 436
Db 2457 ACCCCCAAACTCTAACCCCAATTTCTCT--CCAACCTTACTACCCGAATAAAAAATTA 2400

QY 437 AACAGATGCCAAATCTCCAACA 460
Db 2399 AATAACACCAATATACCAATA 2376

RESULT 13
US-10-311-455-666/c
; Sequence 666, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 666
; LENGTH: 6222
```

```
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-666

Query Match
Best Local Similarity 7.1%; Score 36.4; DB 15; Length 6222;
Matches 79; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 344 AATATCAATATCAGTGGCAGCTTAAGTCTCCAGACCTTACACTTCTTAAGCAATCTCTCTTA 403
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5664 AATTATACCTAATTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 5605
QY 404 CTCACCTTCTCCAGACAGCAAGCAATACAAACAGATGCCAAATCTCCAACTA 463
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5604 CTTTATATCTTTTCACTATAATAATCAACAATACATACATATATACTAAACCCCTA 5545
QY 464 CTGCGACTACCCAAATCTCCACTGATATT 493
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5544 TAAATCCTTACTAAAAATCAACAATCTAT 5515
```

```
RESULT 14
US-10-027-632-135035
; Sequence 135035, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 135035
; LENGTH: 777
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-135035
```

```
Query Match
Best Local Similarity 7.1%; Score 36.2; DB 13; Length 777;
Matches 59; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 371 CTCGAGACCTTACACTTCTTAAGCAATCTCTTACTTCACTTCCAGAACAGCAA 430
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
322 CTCGAGCTGGGCAAGTCAGAGACTTCCATCTTAACCAACACCACCACCAACATAAA 381
QY 431 ATACAAAAACAGATGCCAAAATCTCCACACTACTGC 467
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
382 AAACAAAAACAAACAAACAAAACCTTCCCTCTTTC 418
```

```
RESULT 15
US-10-027-632-135036
; Sequence 135036, Application US/10027632
; Publication No. US20020198371A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 135036
; LENGTH: 777
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-135036

Query Match
Best Local Similarity 7.1%; Score 36.2; DB 13; Length 777;
Matches 59; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 371 CTCGAGACCTTACACTTCTTAAGCAATCTCTTACTTCACTTCCAGAACAGCAA 430
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
322 CTCGAGCTGGGCAAGTCAGAGACTTCCATCTTAACCAACACCACCACCAACATAAA 381
QY 431 ATACAAAAACAGATGCCAAAATCTCCACACTACTGC 467
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
382 AAACAAAAACAAACAAACAAAACCTTCCCTCTTTC 418

Search completed: July 30, 2004, 20:26:37
Job time : 413 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 17:06:36 ; Search time 2908 Seconds

(without alignments)

5237.177 Million cell updates/sec

Title: US-09-386-850-7\_COPY\_1\_510

Perfect score: 510

Sequence: 1 AAACGACTGACGAGAGC.....TTTTCAGGTCGTCGCAA 510

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estnu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_hic:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_hic:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estcom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_pig:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	472.8	92.7	483	10	BF562410
2	399.8	78.4	562	10	AW533278
3	390.8	76.6	402	14	CB769246
4	152.8	30.0	634	10	BF534987

## ALIGNMENTS

RESULT 1  
BF562410  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT

BF562410  
UI-R-BUO-anf-a-06-0-UI.r1 UI-R-BUO Rattus norvegicus cDNA clone  
UI-R-BUO-anf-a-06-0-UI 5', mRNA sequence.  
BF562410.1 GI:11672140  
EST.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Rattus norvegicus  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 483)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
97044477  
889548  
Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu

697 12 BG872108  
581 10 BF539002  
605 10 BF540223  
644 10 BF537882  
652 12 BG865810  
676 10 BF299863  
872 10 BF534149  
652 10 BF302989  
645 10 BF540321  
655 10 BF540217  
653 10 BF536022  
481 10 BF534360  
513 12 BG872364  
745 10 BF302398  
515 10 BF299995  
647 10 BF535583  
649 10 BF538163  
670 10 BF535602  
674 10 BF537472  
867 10 BF540227  
624 10 BF300221  
644 12 BG865669  
728 10 BF300831  
942 12 BG866578  
660 10 BF537022  
942 10 BF302832  
553 12 BG869030  
685 10 BF535415  
645 10 BF300735  
648 10 BF536686  
648 12 BG872654  
650 12 BG870684  
655 10 BF302080  
655 10 BF537445  
660 12 BG872510  
661 11 BG035428  
663 10 BF539051  
664 10 BF538737  
668 12 BG865577  
690 10 BF302310  
695 10 BF540179

BG872108 602792936  
BF539002 602048549  
BF540223 602050271  
BF537882 602049321  
BG865810 602798210  
BF299863 602030493  
BF534149 602047633  
BF302989 602030914  
BF540321 602052377  
BF540217 602050262  
BF536022 602054151  
BF534360 602047860  
BG872364 602792651  
BF302398 602031395  
BF299995 602030654  
BF535583 602051874  
BF538163 602053634  
BF535602 602051893  
BF537472 602053115  
BF540227 602050275  
BF300221 60203126  
BG865669 602783834  
BF300831 602028829  
BG866578 602785548  
BF537022 602048926  
BF302832 602032723  
BG869030 602784661  
BF535415 602051809  
BF300735 602031918  
BF536686 602048662  
BG872654 602793712  
BF302080 602033029  
BF537445 602050084  
BG872510 602793545  
BG035428 Mus muscu  
BF539051 602051406  
BF538737 602051075  
BG865577 602783725  
BF302310 602031281  
BF540179 602050223

cdNA Library Preparation: M.B. Soares Lab Clone distribution:  
clones will be available through Research Genetics (www.resgen.com)  
This clone is also available through the I.M.A.G.E. Consortium at  
LMML (info@lmml.gov). IMAGE ID= 1801619  
Seq primer: M13 Forward.

# FEATURES

Location/Qualifiers  
1. 483  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-BU0-anf-a-06-0-UI"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="UI-R-BU0"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not 1; Site 2: Eco RI; The UI-R-BU0  
library is a subtracted library derived from a mixture of  
eye and ganglia tissues. For a detailed description of the  
library from which this clone was derived, please visit  
our web site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has  
been previously described in (Bonaldo, Lennon and Soares,  
Genome Research 6:791-806, 1996)"

# ORIGIN

Query Match 92.7%; Score 472.8; DB 10; Length 483;  
Best Local Similarity 99.6%; Pred. No. 1.1e-111; Indels 0; Gaps 0;  
Matches 474; Conservative 0; Mismatches 2;

QY 23 CTGACCAGCACATTTCCCGCTCAGAGTTTCTCCAAAGGGGTACCAAGATCAAGTCAC 82  
DB 8 CTGACCAGCACATTTCCCGCTCAGAGTTTCTCCAAAGGGGTACCAAGATCAAGTCAC 67  
QY 83 TGTATTGATCTTTGGCTGTGGATCTTCTAGCATGCTTCCAGTCAGTGGGTGTCA 142  
DB 68 TGTATTGATCTTTGGCTGTGGATCTTCTAGCATGCTTCCAGTCAGTGGGTGTCA 127  
QY 143 GAGGCCCAAGAGACACATATCTTAGAAGACACAAAGATCTTCAACTCTTCTCTCAT 202  
DB 128 GAGGCCCAAGAGACACATATCTTAGAAGACACAAAGATCTTCAACTCTTCTCTCAT 187  
QY 203 ATCTTGGTCTTCAGCTGTATCCCAATGGTGGACAAATAGGAGTAACAATCACTATACCT 262  
DB 188 ATCTTGGTCTTCAGCTGTATCCCAATGGTGGACAAATAGGAGTAACAATCACTATACCT 247  
QY 263 TAAATCTTCAACCACTCTGTCTTCTTAACTTCCCGGTTTATCACTGGACCAACCAT 322  
DB 248 TAAATCTTCAACCACTCTGTCTTCTTAACTTCCCGGTTTATCACTGGACCAACCAT 307  
QY 323 TGGTTGTACAAGTACCACTGAATATCAATATCACTGGCAGCTAACTGCTCCAGACCCCTA 382  
DB 308 TGGTTGTACAAGTACCACTGAATATCAATATCACTGGCAGCTAACTGCTCCAGACCCCTA 367  
QY 383 CACCTCTAAGCAATCTCTTACTCAACTTCTTCCACAGAACAGCAAAATACAAAACAG 442  
DB 368 CACCTCTAAGCAATCTCTTACTCAACTTCTTCCACAGAACAGCAAAATACAAAACAG 427  
QY 443 ATGCCAAATCTCCACACTCTGACACTACCCAAATTCCTCTGATATTTTGA 498  
DB 428 ATGCCAAATCTCCACACTCTGACACTACCCAAATTCCTCTGATATTTTGA 483

# RESULT 2

AW533278/c  
LOCUS  
DEFINITION  
UI-R-BU0-anf-a-06-0-UI.s1 UI-R-BU0 Rattus norvegicus cDNA clone  
UI-R-BU0-anf-a-06-0-UI 3', mRNA sequence.  
AW533278  
ACCESSION  
VERSION  
AW533278.1 GI:7175692  
KEYWORDS  
Rattus norvegicus (Norway rat)  
SOURCE  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

# REFERENCE

AUTHORS  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE  
Normalization and subtraction: two approaches to facilitate gene  
discovery

# JOURNAL

MEEDLINE  
97044477  
PUBMED  
889548

# COMMENT

Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565

Email: [bento-soares@uiowa.edu](mailto:bento-soares@uiowa.edu)

The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to identify it as a clone from the  
normalized ganglia library cDNA library Preparation: M.B. Soares  
Lab Clone distribution: clones will be available through Research  
Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA=Yes.

# FEATURES

Location/Qualifiers  
1. 562  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-BU0-anf-a-06-0-UI"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="UI-R-BU0"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not 1; Site 2: Eco RI; The UI-R-BU0  
library is a subtracted library derived from a mixture of  
eye and ganglia tissues. For a detailed description of the  
library from which this clone was derived, please visit  
our web site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has  
been previously described in (Bonaldo, Lennon and Soares,  
Genome Research 6:791-806, 1996)

TAG\_TISSUE=ganglia

TAG\_LIB=UI-R-BU0

TAG\_SEQ=GCAGAC

# ORIGIN

Query Match 78.4%; Score 399.8; DB 10; Length 562;  
Best Local Similarity 99.5%; Pred. No. 8.9e-93;  
Matches 401; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 108 CCTTCTAGCATCTTCCAGTCAGGTGAGGTCTCAGAGCCCAAGACACATAATCC 167  
DB 562 CCTTCTAGCATCTTCCAGTCAGGTGAGGTCTCAGAGCCCAAGACACATAATCC 503  
QY 168 TAGAGACAAACAGATCCTTCAACTCTTCTCATTTATCTTGGTCTTCAGCCTGATCCAA 227  
DB 502 TAGAGCCCAACAGATCCTTCAACTCTTCTCATTTATCTTGGTCTTCAGCCTGATCCAA 443  
QY 228 TGGTGACAAATAGGAGTAACAATCACTATACCTTAAATCTTCAACACCTCGTGTCT 287  
DB 442 TGGTGACAAATAGGAGTAACAATCACTATACCTTAAATCTTCAACACCTCGTGTCT 383  
QY 288 TGTATATCTTCCCGTTTATCACTGGACCACTTGTGTACAGGTACCACTCAATA 347  
DB 382 TGTATATCTTCCCGTTTATCACTGGACCACTTGTGTGTACAGGTACCACTCAATA 323  
QY 348 TCAATATCAGTGGCAGCTAACTGCTCCAGACCCCTACACCTCTAAGCAATCTCTCTACTCA 407  
DB 322 TCAATATCAGTGGCAGCTAACTGCTCCAGACCCCTACACCTCTAAGCAATCTCTCTACTCA 263

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus

1 (bases 1 to 562)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene  
discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: [bento-soares@uiowa.edu](mailto:bento-soares@uiowa.edu)



```
QY 408 ACTTATTCCACAGAACAAATACAAACAGATGCCAAATCTCCAACTACTGC 467
|||||
Db 262 ACTTCTTCCACAGAACAAATACAAACAGATGCCAAATCTCCAACTACTGC 203
|||||
QY 468 GACTACCCAAAATTCCTGATATTTTGAAGGTGGTGGCAA 510
|||||
Db 202 GACTACCCAAAATTCCTGATATTTTGAAGGTGGTGGCAA 160
|||||

RESULT 3
CB769246 402 bp mRNA linear EST 16-MAY-2003
LOCUS AMGNNUC:SRPB2-00151-E10-A.srbp2 (10220) Rattus norvegicus cDNA
DEFINITION clone srbp2-00151-e10 5', mRNA sequence.
ACCESSION CB769246
VERSION CB769246.1 GI:29857637
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
REFERENCE 1 (bases 1 to 402)
AUTHORS Angen EST Program.
TITLE Angen Rat EST Program
JOURNAL Unpublished (2003)
COMMENT Contact: Dan Fitzpatrick
Angen, Inc
One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00151 row: e column: 10.
FEATURES
Location/Qualifiers
1..402
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="srbp2-00151-e10"
/tissue_type="prostate tissue"
/clone_lib="srbp2 (10220)"
/notes="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; rat
prostate normalized double selected poly(A+) mRNA size
fraction > 1 kb"
ORIGIN
Query Match 76.6%; Score 390.8; DB 14; Length 402;
Best Local Similarity 99.5%; Pred. No. 1.8e-90;
Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 4 CTGACTGACGAGAGCTTCTGACACGACATTTCCCGCTCAGAAGTTTCTCCAGGGG 63
Db 9 CCGACTGACGAGAGCTTCTGACACGACATTTCCCGCTCAGAAGTTTCTCCAGGGG 68
QY 64 CTACCAAGATGAAGTCACTGATTTGATCTTTGGCTGTGGATCTCTAGCATCTTC 123
Db 69 CTACCAAGATGAAGTCACTGATTTGATCTTTGGCTGTGGATCTCTAGCATCTTC 128
QY 124 CAGTCAGGTGAGGTGTCAGAGCCCAAGACACATATCTTACAGACACAAAGAT 183
Db 129 CAGTCAGGTGAGGTGTCAGAGCCCAAGACACATATCTTACAGACACAAAGAT 188
QY 184 CTTTCAACTCTTCCCTATATCTTTGCTTTCAGCTGATCCCAATGGTGACAAATAGGA 243
Db 189 CTTTCAACTCTTCCCTATATCTTTGCTTTCAGCTGATCCCAATGGTGACAAATAGGA 248
QY 244 GTAACATCACTATACCTTAAATCTTCAACCACTCGTGTGTTCTTGTAAATCTTCCCGGT 303
Db 249 GTAACATCACTATACCTTAAATCTTCAACCACTCGTGTGTTCTTGTAAATCTTCCCGGT 308
QY 304 TTTATCACTGGACCACTGTTGTACAGGTACCACTGAATATCAATATCAGTGGCAG 363
Db 309 TTTATCACTGGACCACTGTTGTACAGGTACCACTGAATATCAATATCAGTGGCAG 368
QY 364 CTAAGTCTCCAGACCTTACCTCTTAAGCAATC 397
```

```
Db 369 CTAAGTCTCCAGACCTTACACCTTAAAGCAATC 402
|||||
```

```
RESULT 4
BF534987 634 bp mRNA linear EST 11-DEC-2000
LOCUS 602050440F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4190048 5',
DEFINITION mRNA sequence.
ACCESSION BF534987
VERSION BF534987.1 GI:11622350
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 634)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9516 row: o column: 09
High quality sequence stop: 614.
FEATURES
Location/Qualifiers
1..634
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4190048"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP SG2"
/notes="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:
NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
ORIGIN
Query Match 30.0%; Score 152.8; DB 10; Length 634;
Best Local Similarity 62.4%; Pred. No. 8.3e-29;
Matches 318; Conservative 0; Mismatches 167; Indels 25; Gaps 4;
QY 1 AAACCTGACTGACGAGAGCTTCTGACACGACATTTCCCGCTCAGAAGTTTCTCCAAG 60
Db 5 AAAATAACTGACGAGAGCTTCTGACACGACATTTCCCGCTCAGAAGTTTCTCCAAG 64
QY 61 GGGCT-ACCAAGATGAAGTCACTGATTTGATCTTTGGCTGTGGATCTCTAGCATG 119
Db 65 AGGCTGAAAAAAGATGAAGCCCAATTTGCTTTAGGCTCTGCTTTGTTAGTAGT 124
QY 120 CTTTCAGTCAAGTGAAGGTGTCAGAGGCCCAAGACACATATCTTAGAGACACA 179
Db 125 CTTTCGTGCTGTGAGTGTACAGAGGCCCTTAGAAG---ACATGATCTTAGAGCAAT 181
QY 180 AGATCCTTTCAATCTTCTCTCATTTATCTTTGGCTTTCAGCTGATCCCAATGGTGGACAAT 239
Db 182 CCTCTCTCTCTCCCTCCCTCCATGGTCCAG-----AATTGGTAGACCACA 226
QY 240 AGGAGTAAATCACTATACCTTAAATCTTCAACCACTCGTGTGTTGTTTAAATCTTCC 299
Db 227 CCCTCCACCTTTTGGTCCAGGAATTTGGTAGACCACTCCCTTTGGTCCAGGAAT 286
QY 300 CGGTTTTATCTCTGACCACTGATTTGGTTGTACAAGGTACCACTGAATATCAATATCAGTG 359
Db 287 TGGTCGAAGCGCCCTTCCACTCTCTTTGTCACCACTTCTTCCACATCTTAGACCTCCAAG 346
```

QY 360 GCAGTAAGTCTCCAGACCCCTACACCTCTAAGCAATCCTCTACTCACTTCACTCCAC 419  
 DB 347 CAATCCGCTCTCTCCA-----CTTACTCCAGCAATCTCTCTACCGGACCTCTTACCAC 400  
 QY 420 AGAACAGCAATACAAAAAGAGTCCAAATCTCCAAACACTACTGCGCACTACCCAAA 479  
 DB 401 AGTACAAGCACTACAAATGCCAGCTGCCAGTATCTCCATACTACTCTCTACTGCTAGAGA 460  
 QY 480 TTCCACTGATATTTTGAAGTGGTGGCAA 509  
 DB 461 TTCCACTGATATTTTGGAGGCTGGGAA 490

RESULT 5  
 BG872108 602792936F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:4924089 5',  
 LOCUS mRNA sequence.

ACCESSION BG872108  
 VERSION BG872108.1 GI:14222648

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 697)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM10846 row: h column: 10

High quality sequence stop: 635.

Location/Qualifiers

FEATURES

source

1..697

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:4924089"

/lab\_host="DH10B (T1 phage-resistant)"

/clone\_lib="NCI\_CGAP\_SG2"

/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site: 1:

NotI; Site: 2: SalI; Cloned unidirectionally. Primer: Oligo

dr. Average insert size 1.3 Kb. Constructed by Life

Technologies. Note: this is a NCI\_CGAP Library."

ORIGIN

Query Match 29.8%; Score 152.2; DB 12; Length 697;

Best Local Similarity 62.3%; Pred. No. 1.2e-28;

Matches 317; Conservative 0; Mismatches 168; Indels 24; Gaps 4;

QY 2 AACTGACTGACAGAGCTTGTACAGACATTTCCCGCTCAGAGTTTCTCAAG 61

DB 1 AAATACTGACAGACCGCTTGTACAGACATTTCCAGCTAAGAGTGTCTCCAAG 60

QY 62 GGCT-ACCAAGATGAAGTCACTGTATTGATCTTTGGCTGTGGATCTCTAGATGC 120

DB 61 GGCTGAAAAGATGAAGCCACTCAATTTGGTCTTAGGCCTCTGSCATTTGTAGGATGC 120

QY 121 TTCAGTCAAGTGAAGGTGTACAGAGCCCAAGAGCAACATAATCTTAGAGCAACAA 180

DB 121 TTCTGTCTGTGAGTGTACAGAGCCCTTAGAG--ACATGATCTAGAGCCATTC 177

QY 181 GATCCTTCAACTCTTCTCTATATCTTGGTCTTTCAGCCTGATCCCAATGTGCAATA 240

DB 178 CCTCTCTCTCCCTCCCATGGTCCAGGAATT---GGTAGACCACACCCCTCCACCTTT 234  
 QY 241 GGAGTAACAATCACTATACCCCTTAAATCTTCAACACCTCGTGTCTTGTAACTTCCC 300  
 DB 235 GGTCCAGGAATGGTAGACACACCCCTCCACCTTTGGTCCAGGAATTTGGTCCGACCCC 294  
 QY 301 GGTTTTATCACTGGACACCACTTGGTTCTTACAAGGTACCACTGAATATCAATATCAGTGG 360  
 DB 295 CCTCAACTCTTGTCCAAAGATTCTCTCCACATCTAGACCT-----C 337  
 QY 361 CAGCTAACTGCTCCAGACCTTACACCTCTAAGCAATCTCTCTACTCAACTTCACTCCACA 420  
 DB 338 CAAGCAATCCGTCTCTCCATCCACTACTCCAAGCATCTCTCTACCGACCTCTTACCACA 397  
 QY 421 GAACAGCAATACAAAAAGAGTCCAAATCTCCAACTACTGCGACTACCCAAAT 480  
 DB 398 GTACAAGCAACTACAAATGCCAGTCCAGTATCTCCATACTACTCTCTACTCTAGAT 457  
 QY 481 TCCACTGATATTTTGAAGTGGTGGCAA 509  
 DB 458 TCCACTGATATTTTGGAGTTGTGGAA 486

RESULT 6

BF539002

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 581)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM9511 row: a column: 14

High quality sequence stop: 580.

Location/Qualifiers

FEATURES

source

1..581

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:4187797"

/lab\_host="DH10B (T1 phage-resistant)"

/clone\_lib="NCI\_CGAP\_SG2"

/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site: 1:

NotI; Site: 2: SalI; Cloned unidirectionally. Primer: Oligo

dr. Average insert size 1.3 kb. Constructed by Life

Technologies. Note: this is a NCI\_CGAP Library."

ORIGIN

Query Match 29.8%; Score 151.2; DB 10; Length 581;

Best Local Similarity 62.2%; Pred. No. 2.1e-28;

Matches 317; Conservative 0; Mismatches 168; Indels 25; Gaps 4;

QY 1 AAATCACTGACAGAGAGCTTCTGACAGACATTTCCCGCTCAGAGTTTCTCCAAG 60

DB 31 AAATACTGACAGACCGCTTCTGACAGACATTTCCAGCTAAGAGTGTCTCAAG 90



/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:418749"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NCI CGAP SG2"  
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:  
Note; Site 2: Sal; Cloned unidirectionally. Primer: Oligo  
dt. Average insert size 1.3 kb. Constructed by Life  
Technologies. Note: this is a NCI CGAP Library."

## ORIGIN

Query Match 29.6%; Score 151.2; DB 10; Length 644;  
Best Local Similarity 62.2%; Pred. No. 2.2e-28;  
Matches 317; Conservative 0; Mismatches 168; Indels 25; Gaps 4;  
  
QY 1 AAATGACTGACGACGAGCTCTTCACACGACATTTCCCGCTCAGAGTTTCTCCAAAG 60  
Db 3 AAATTAATCTGACGACGCTCTTCACACGACATTCACGCTAAGAGTGTCTCCAAAG 62  
  
QY 61 GGGCT-ACCAAGATGAAGTCACTGTATTGTATCTTTGGCTGTGGATCCTTCTAGCATG 119  
Db 63 AGGCTGAAAAAAGATGAAGCCACTCAATTTGGTCTTAGGCTCTGCAATCTTTAGGATG 122  
  
QY 120 CTTCCAGTCAGTGAGGTTGTGAGAGGCCCAAGACACATATTCCTAGAGACAACA 179  
Db 123 CTTCTCTGCTGTGAGTGTGACAGAGGCCCTAGAAG---ACATGATCCTAGAGGACCAT 179  
  
QY 180 AGATCCTTCAACTCTTCTCTCATATCTTGTCTTTCAGCCTGATCCCAATGGTGGACAAT 239  
Db 180 CCTCTCTCTCCCTCCCTCCCTGCTCAGGAAT---GGTAGACACACCCCTCCACCTT 236  
  
QY 240 AGGAGTAACAATCACTATACCTTAAATCTTCAACACCTGTGTCTTTGTTAACTTCC 299  
Db 237 TGGTCCAGGAATTTGGTAGACACCCCTCCACCTTTGGTCCAGGAATTTGGTCCAGCAC 296  
  
QY 300 CGGTTTATCACTGGACACCATTTGTTGTAGAGGTACCCTGATCAATATCAATATCAGTG 359  
Db 297 CCTCCACCTCTCTGTGTCACCACTTCTCCACATCTTAGACCT----- 339  
  
QY 360 GCAGCTAACTGTCTCCAGACCTTACACTCTTAAGCAATCTCTTACTCAACTTCCAC 419  
Db 340 -CCAAGCAATCGTCTCTCCAGTACTCCAAGCTTCTCTTACCGGACCTCTTACCAC 398  
  
QY 420 AGAACAGCAATACAAAACAGATGCCAAATCTCCACACTACTGCGACTACCCAAA 479  
Db 399 AGTACAGCAACTACAAATGCCAGTCCAGTATCTCCATAACTACTCTACTCTAGAGA 458  
  
QY 480 TTCCTACTGATATTTTGAAGGTGGTGGCA 509  
Db 459 TTCCTACTGATATTTTGGAGTTGTGGAA 488

## RESULT 9

BG65810  
LOCUS  
DEFINITION  
602788210F1 NCI CGAP SG2 Mus musculus cDNA clone IMAGE:4914002 5',  
mRNA sequence.  
ACCESSION  
BG65810  
VERSION  
BG65810.1 GI:14216350  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 652)  
AUTHORS  
NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
Unpublished (1999)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM10820 row: d column: 03  
High quality sequence stop: 636.  
Location/Qualifiers

## FEATURES

source  
1..652  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4914002"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NCI CGAP SG2"  
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:  
Note; Site 2: Sal; Cloned unidirectionally. Primer: Oligo  
dt. Average insert size 1.3 kb. Constructed by Life  
Technologies. Note: this is a NCI CGAP Library."

## ORIGIN

Query Match 29.6%; Score 151.2; DB 12; Length 652;  
Best Local Similarity 62.2%; Pred. No. 2.2e-28;  
Matches 317; Conservative 0; Mismatches 168; Indels 25; Gaps 4;  
  
QY 1 AAATGACTGACGACGAGCTTCTGACGACATTTCCCGCTCAGAGTTTCTCCAAAG 60  
Db 3 AAATTAATCTGACGACGCTTCTGACGACATTCACGCTAAGAGTGTCTCCAAAG 62  
  
QY 61 GGGCT-ACCAAGATGAAGTCACTGTATTGTATCTTTGGCTGTGGATCCTTCTAGCATG 119  
Db 63 AGGCTGAAAAAAGATGAAGCCACTCAATTTGGTCTTAGGCTCTGCAATCTTTAGGATG 122  
  
QY 120 CTTCCAGTCAGTGAGGTTGTGAGAGGCCCAAGACACATATTCCTAGAGACAACA 179  
Db 123 CTTCTCTGCTGTGAGTGTGACAGAGGCCCTAGAAG---ACATGATCCTAGAGGACCAT 179  
  
QY 180 AGATCCTTCAACTCTTCTCTCATATCTTGTCTTTCAGCCTGATCCCAATGGTGGACAAT 239  
Db 180 CCTCTCTCTCCCTCCCTCCCTGCTCAGGAAT---GGTAGACACACCCCTCCACCTT 236  
  
QY 240 AGGAGTAACAATCACTATACCTTAAATCTTCAACACCTGTGTCTTTGTTAACTTCC 299  
Db 237 TGGTCCAGGAATTTGGTAGACACCCCTCCACCTTTGGTCCAGGAATTTGGTCCAGCAC 296  
  
QY 300 CGGTTTATCACTGGACACCATTTGTTGTAGAGGTACCCTGATCAATATCAATATCAGTG 359  
Db 297 CCTCCACCTCTCTGTGTCACCACTTCTCCACATCTTAGACCT----- 339  
  
QY 360 GCAGCTAACTGTCTCCAGACCTTACACTCTTAAGCAATCTCTTACTCAACTTCCAC 419  
Db 340 -CCAAGCAATCGTCTCTCCAGTACTCCAAGCTTCTCTTACCGGACCTCTTACCAC 398  
  
QY 420 AGAACAGCAATACAAAACAGATGCCAAATCTCCACACTACTGCGACTACCCAAA 479  
Db 399 AGTACAGCAACTACAAATGCCAGTCCAGTATCTCCATAACTACTCTACTCTAGAGA 458  
  
QY 480 TTCCTACTGATATTTTGAAGGTGGTGGCA 509  
Db 459 TTCCTACTGATATTTTGGAGTTGTGGAA 488

## RESULT 10

BF299863  
LOCUS  
DEFINITION  
602030493F1 NCI CGAP SG2 Mus musculus cDNA clone IMAGE:4165726 5',  
mRNA sequence.  
ACCESSION  
BF299863  
VERSION  
BF299863.1 GI:11246386  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)

ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 676)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1AM9453 row: i column: 23  
High quality sequence stop: 626.  
Location/Qualifiers  
1..676  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4165726"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NCI CGAP SG2"  
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:  
NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo  
dT. Average insert size 1.3 kb. Constructed by Life  
Technologies. Note: this is a NCI CGAP Library."

FEATURES  
source  
1..676  
Query Match 29.6%; Score 151.2; DB 10; Length 676;  
Best Local Similarity 62.2%; Pred. No. 2.2e-28;  
Matches 317; Conservative 0; Mismatches 168; Indels 25; Gaps 4;  
QY 1 AAATGACTGACGACAGAGCTTCTGACGACGACATTTCCCGCTCAGAGTTTCTCCAAG 60  
Db 1 AAATAACTGACGACAGCGCTTCTGACGACGACATTCACGAGTAAGAGTGTCTCCAAG 60  
QY 61 GGGCT-ACCAAAGATGAAGTCACTGTATTTGATCTTTGGCTGTGGATCCTCTAGCATG 119  
Db 61 AGGCTGAAAAGATGAAGCCACTCAATTTGGTCTTAGGCTCTGCATCTTGTAGATG 120  
QY 120 CTTCCAGTCAGTGAGGGTGTGAGAGCCCAAGAGCAACATAATCTTAGAGAGCAACA 179  
Db 121 CTTCCCTGCTGTGAGTGTACAGAGGCCCTTAGAG---ACATGATCCTAGAGGACCAT 177  
QY 180 AGATCCTTCAACTCTTCTCATTTATCTTGTCTTTCAGCCTGATCCCAATGGTGGACAA 239  
Db 178 CCCTCTCTCTCCCTTCCCGATGTCAGGAAT---GGTAGACACACCCCTCCACCTT 234  
QY 240 AGAGTAAACAATCACTATACCTTAAATCTTCAACACCTCTGTGTTCTTTAATCTTCC 299  
Db 235 TGGTCCAGGAATGTGAGACACCCCTTCCACCTTTGGTCAGGAATGGTTCGACCCACC 294  
QY 300 CGGTTTTATCACTGGACCAACCATTTGGTGTACAGGTACCACTGCACTACCCAAAA 359  
Db 295 CCCTCCACCTCTCTGTCACCAAGTCTTCCCACTCCTAGACCT----- 337  
QY 360 GCAGCTAACTGCTCCAGACCCCTACCTCTTAAGCAATCTCTACTCAACTTCATTCAC 419  
Db 338 -CCAAGCAATCGCTCTCTCCACTACTCAAGCAATCTCTTACCGGACCTCTTACCAC 396  
QY 420 AGAACAAGCAATACAAAACAGATGCCAAATCTCCAACTACTGCACTACCCAAAA 479  
Db 397 AGTAAAGCAACTACAATGCCAGTCCAGTATCTCCATACTACTCTTACTGCTAGAG 456  
QY 480 TTCCACTGATATTTTGAAGGTGGGCAA 509  
Db 457 TTCCACTGATATTTTGGAGGTGTGGGAA 486

RESULT 11  
LOCUS BF534149  
DEFINITION 602047633F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:4187024 5',  
mRNA sequence.  
ACCESSION BF534149  
VERSION BF534149.1 GI:11621512  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 872)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1AM9509 row: a column: 09  
High quality sequence stop: 621.  
Location/Qualifiers  
1..872  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4187024"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NCI CGAP SG2"  
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:  
NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo  
dT. Average insert size 1.3 kb. Constructed by Life  
Technologies. Note: this is a NCI CGAP Library."

ORIGIN  
Query Match 29.3%; Score 149.6; DB 10; Length 872;  
Best Local Similarity 62.0%; Pred. No. 5.9e-28;  
Matches 316; Conservative 0; Mismatches 169; Indels 25; Gaps 4;  
QY 1 AAATGACTGACGACAGAGCTTCTGACGACGACATTTCCCGCTCAGAGTTTCTCCAAG 60  
Db 1 AAATAACTGACGACAGCGCTTCTGACGACGACATTCACGAGTAAGAGTGTCTCCAAG 60  
QY 61 GGGCT-ACCAAAGATGAAGTCACTGTATTTGATCTTTGGCTGTGGATCCTCTAGCATG 119  
Db 61 AGGCTGAAAAGATGAAGCCACTCAATTTGGTCTTAGGCTCTGCATCTTGTAGATG 120  
QY 120 CTTCCAGTCAGTGAGGGTGTGAGAGCCCAAGAGCAACATAATCTTAGAGAGCAACA 179  
Db 121 CTTCCCTGCTGTGAGTGTACAGAGGCCCTTAGAG---ACATGATCCTAGAGGACCAT 177  
QY 180 AGATCCTTCAACTCTTCTCATTTATCTTGTGTTTTCAGCCTGATCCCAATGGTGGACAA 239  
Db 178 CCCTCTCTCTCCCTTCCCGATGTCAGGAAT---GGTAGACACACCCCTCCACCTT 234  
QY 240 AGAGTAAACAATCACTATACCTTAAATCTTCAACACCTCTGTGTTCTTTAATCTTCC 299  
Db 235 TGGTCCAGGAATGTGAGACACCCCTTCCACCTTTGGTCAGGAATGGTTCGACCCACC 294  
QY 300 CGGTTTTATCACTGGACCAACCATTTGGTGTACAGGTACCACTGCAATATCAATATCAG 359  
Db 295 CCCTCCACCTCTCTGTCACCAAGTCTTCCCACTCCTAGACCT----- 337  
QY 360 GCAGCTAACTGCTCCAGACCCCTACCTCTTAAGCAATCTCTACTCAACTTCATTCAC 419

Db 338 -CCAGCAATCGTCTCTCCACCTACTCCCAAGCATTTCTCTCCAGGAGCTCTTACCAC 396  
Qy 420 AGAACAGCAATACAAACACGATGCGCAAAATCTCCACACTACTCGGACTACCCAAA 479  
Db 397 AGTACAAGCACTACAATGCCAGCTGCCAGTATCTCCATACTACTCTCTACTGTAGAGA 456  
Qy 480 TTCCACTGATATTTTGAAGGTGGGCAA 509  
Db 457 TTCCACTGATATTTTGGAGGTGGGAA 486

RESULT 12  
BF302989  
LOCUS 602030914F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:4166139 5',  
DEFINITION mRNA sequence.

ACCESSION BF302989  
VERSION BF302989  
KEYWORDS EST.

SOURCE Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 652)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-rc@mail.nih.gov](mailto:cgapbs-rc@mail.nih.gov)

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM9454 row: k column: 04

High quality sequence stop: 645.

## FEATURES

source

1..652

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:4166139"

/lab\_host="NCI CGAP\_SG2"

/clone\_lib="NCI (T1 phage-resistant)"

/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:

NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo

dt. Average insert size 1.3 kb. Constructed by Life

Technologies. Note: this is a NCI\_CGAP Library."

## ORIGIN

Query Match 29.3%; Score 149.2; DB 10; Length 652;  
Best Local Similarity 61.8%; Pred. No. 7.2e-28;  
Matches 315; Conservative 0; Mismatches 170; Indels 25; Gaps 4;

Qy 1 AAATCACTGACAGAGAGCTTCTGACAGACATTTCCCGCTCAGAGTCTTCCAAAG 60  
Db 5 AAATAACTGACAGACCGCTTCTGACAGACATTCACCAAGTAAAGTGTCTCCAAG 64

Qy 61 GGCGT-ACCAAGATGAAGTCACTGTATTGTCTTTGGCTGTGGATCCTCTAGCATG 119  
Db 65 AGSCTGAARAAGATGAGCCACTCAATTGGTCTTAGGCTCTGCATTCTTGTAGATG 124

Qy 120 CTTCCAGTCAAGGTGTGAGAGCCCAAGACACATTAATCTTAGAGACACACA 179  
Db 125 CTTCCCTGCTGTGAGTGTACAGAGGCCCTAGAG---ACATGATCTTAGAGGCCATT 181

Qy 180 AGATCCTTCAACTCTTCTCATTTATCTTGGTCTTCAGCCCTGATCCCAATGGTGAACAA 239  
Db 182 CCCTCTCTCTCCCTCTCCCATGTTCAGGAANT--GGTAGACCACACCTCCACCTT 238

Qy 240 AGGAGTAACAATCACTATACCTTAAATCTTCAACCACTCGTGTGTGTAAATCTTCC 299  
Db 239 TGGTCCAGGAATGGTAGACACCCCTCCACCTGTGGTCCAGGAATTGGTCCGACC 298  
Qy 300 CGGTTTATCACTGACCAACCATTTGGTTGTACAGGTACCACTGAATATCAATATCAGTG 359  
Db 299 CCTCCACCTCTGTGTCCACCAAGTTCTCTCCATCTCTAGACCT----- 341  
Qy 360 GCAGCTAACTGCTCCAGACCTTACACCTCTTAAGCAATCTCTACTCAACTTCATTCCAC 419  
Db 342 -CCAAGCAATCGTCTCTCCACTACTCCAGCAATTCTCTACCGGACCTCTTACCAC 400  
Qy 420 AGAACAGCAATACAAACACAGATGCCAAATCTCCAACTACTCGGACTACCCAAA 479  
Db 401 AGTACAAGCACTACAATGCCAGTGCAGTATCTCCATACTACTCTCTACTGTAGAGA 460  
Qy 480 TTCCACTGATATTTTGAAGGTGGGCAA 509  
Db 461 TTCCACTGATATTTTGGAGGTAGTGGAA 490

## RESULT 13

BF540321

LOCUS 602052377F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:4191474 5',

DEFINITION mRNA sequence.

ACCESSION BF540321

VERSION BF540321.1

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 645)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-rc@mail.nih.gov](mailto:cgapbs-rc@mail.nih.gov)

Tissue Procurement: Jeffrey E. Green, M.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM9520 row: j column: 19

High quality sequence stop: 638.

Location/Qualifiers

1..645

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:4191474"

/lab\_host="NCI CGAP\_SG2"

/clone\_lib="NCI (T1 phage-resistant)"

/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:

NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo

dt. Average insert size 1.3 kb. Constructed by Life

Technologies. Note: this is a NCI\_CGAP Library."

Technologies. Note: this is a NCI\_CGAP Library."

## ORIGIN

Query Match 29.0%; Score 148; DB 10; Length 645;  
Best Local Similarity 61.8%; Pred. No. 1.5e-27;  
Matches 315; Conservative 0; Mismatches 170; Indels 25; Gaps 4;

Qy 1 AAATCACTGACAGAGAGCTTCTGACAGACATTTCCCGCTCAGAGTCTTCCAAAG 60  
Db 1 AAATAACTGACAGACCGCTTCTGACAGACATTCACCAAGTAAAGTGTCTCCAAG 60

Qy 61 GGCGT-ACCAAGATGAAGTCACTGTATTGTCTTTGGCTGTGGATCCTCTTAGCATG 119  
Db 65 AGSCTGAARAAGATGAGCCACTCAATTGGTCTTAGGCTCTGCATTCTTGTAGATG 124

Db 61 AGGCTGAAAAAAGATGAAGCCACTCAATTTGGTCTTAGGCTCTGCATCTTTGTAGGATG 120  
QY 120 CTTCCAGTCAAGTCAAGGTGTACAGAGCCGAGAGCAACATAATCTTAGAACAACA 179  
Db 121 CTTCCAGTCAAGTGTACAGAGCCGAGAGCAACATAATCTTAGAACAACA 177  
QY 180 AGATCCCTTCAACTCTCTCTCAATATCTTGGTCTTAGGCTCTGCATCTTTGTAGGATG 239  
Db 178 CCCTCCCT 234  
QY 240 AGGAGTAACTACTATACCTTTAAATCTTCAACACCTCTCTCTCTCTCTCTCTCTCTCTCT 299  
Db 235 TGTCTCCAGGAATTTGGTAGACACCCCTCCACCCCTTTGGTCCAGGAATTTGGTCCAGCAC 294  
QY 300 CGGTTTTATCACTGGACCACTATTTGGTGTACAGAGTACCACTGAATATCAATATCAGTG 359  
Db 295 CCCTCCACCT 337  
QY 360 GCAGCTAACTGCTCCAGAGCCCTACACCTCTAAGCAATCTCTCTCTCTCTCTCTCTCTCTCTCT 419  
Db 338 -CCAAGCAATCCGCT 396  
QY 420 AGAACAAGCAAAATACAAAACAGATGCAAAATCTCCAACTACTCTCGACTACCCCAAAA 479  
Db 397 AGTACAAGCAAACTACTGTCAGCTGCGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 456  
QY 480 TTCCACTGATATTTTGAAGGTGGTGGCAA 509  
Db 457 TTCCACTGATATTTTGAAGGTGGTGGCAA 486

RESULT 14  
BF540217 655 bp mRNA linear EST 11-DEC-2000  
LOCUS 602050262F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:4189599 5',  
DEFINITION mRNA sequence.

ACCESSION BF540217  
VERSION BF540217.1 GI:11627598  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NIH-MGC <http://mgc.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLM9515 row: 1 column: 16  
High quality sequence stop: 637.  
Location/Qualifiers  
1. .655  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone\_lib="IMAGE:4189599"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NCI\_CGAP\_SG2"  
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site: 1:  
NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo  
dT. Average insert size 1.3 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

FEATURES  
source

Query Match 29.0%; Score 148; DB 10; Length 655;  
Best Local Similarity 61.8%; Pred. No. 1.5e-27;  
Matches 315; Conservative 0; Mismatches 170; Indels 25; Gaps 4;  
QY 1 AACTGCTGCTACACAGAGAGCTTCTGACAGACATTTCCCGCTCAGAAAGTTTCTCCAA 60  
Db 5 AATAAATACTACACAGAGAGCTTCTGACAGACATTTCCAGAGTGTCTCCAA 64  
QY 61 GGGCT-ACCAAGATGAAGTCACTGTATTTGATCTTTGGCTGTGGATCTTCTAGCATG 119  
Db 65 AGGCTGAATAAGATGAGCCACTCAATTTGGTCTTAGGCTCTGCATCTTTGTAGGATG 124  
QY 120 CTTCCAGTCAAGGTGTACAGAGCCGAGAGCAACATAATCTTAGAACAACA 179  
Db 125 CTTCCAGTCAAGGTGTACAGAGCCGAGAGCAACATAATCTTAGAACAACA 181  
QY 180 AGATCCCTTCAACTCTCTCTCAATATCTTGGTCTTAGGCTCTGCATCTTTGTAGGATG 239  
Db 182 CCCTCCCT 238  
QY 240 AGGAGTAACTACTATACCTTTAAATCTTCAACACCTCTCTCTCTCTCTCTCTCTCTCTCT 299  
Db 239 TGTCTCCAGGAATTTGGTAGACACCCCTCCACCCCTTTGGTCCAGGAATTTGGTCCAGCAC 298  
QY 300 CGGTTTTATCACTGGACCACTATTTGGTGTACAGAGTACCACTGAATATCAATATCAGTG 359  
Db 299 CCCTCCACCT 341  
QY 360 GCAGCTAACTGCTCCAGAGCCCTACACCTCTAAGCAATCTCTCTCTCTCTCTCTCTCTCTCTCT 419  
Db 342 -CCAAGCAATCCGCT 400  
QY 420 AGAACAAGCAAAATACAAAACAGATGCAAAATCTCCAACTACTCTCGACTACCCCAAAA 479  
Db 401 AGTACAAGCAAACTACTGTCAGCTGCGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 460  
QY 480 TTCCACTGATATTTTGAAGGTGGTGGCAA 509  
Db 461 TTCCACTGATATTTTGAAGGTGGTGGCAA 490

RESULT 15  
BF536022

LOCUS 602054151F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:4193269 5',  
DEFINITION mRNA sequence.

ACCESSION BF536022.1 GI:11623390  
VERSION BF536022  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 653)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLM9525 row: e column: 14  
High quality sequence stop: 653.  
Location/Qualifiers  
1. .653  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"

FEATURES  
source





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 17:10:46 ; Search time 303 Seconds  
(without alignments)  
934.076 Million cell updates/sec

Title: US-09-386-850-7\_COPY\_1\_510

Perfect score: 510

Sequence: 1 AAACGACTGACACAGAGC.....TTTTGAAGGTGGTGCAAA 510

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:\*

1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*

2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*

3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*

4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*

5: /cgn2\_6/ptodata/2/ina/6C COMB.seq.\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	510	100.0	658	3	US-08-476-120-7
2	36	7.1	644	4	Sequence 7, Appli
3	36	7.1	644	4	Sequence 752, App
4	35.8	7.0	832	4	Sequence 752, App
5	35	6.9	864	4	Sequence 2813, Ap
6	35	6.9	7218	1	Sequence 2584, Ap
7	34.2	6.7	1150	4	Sequence 14, Appl
8	34.2	6.7	1150	4	Sequence 69, Appl
9	33.8	6.6	116592	4	Sequence 3, Appli
10	33.6	6.6	5208	3	Sequence 70, Appl
11	33.6	6.6	5208	4	Sequence 11, Appl
12	33.6	6.6	5208	4	Sequence 70, Appl
13	33.6	6.6	1664976	4	Sequence 11, Appl
14	33.4	6.5	528	4	Sequence 70, Appl
15	33	6.5	33	4	Sequence 12, Appl
16	32.8	6.4	399	4	Sequence 8, Appli
17	32.6	6.4	399	4	Sequence 8976, Ap
18	32.6	6.4	169998	4	Sequence 6950, Ap
19	32.6	6.4	197496	4	Sequence 24, Appl
20	32.4	6.4	399	4	Sequence 10, Appl
21	32.4	6.4	425	4	Sequence 15653, A
22	32.2	6.3	1193	4	Sequence 15651, A
23	32.2	6.3	1193	4	Sequence 119, App
24	32.2	6.3	392000	4	Sequence 119, App
25	32.2	6.3	1230025	4	Sequence 11, Appl
26	32	6.3	1333	3	Sequence 1, Appli
27	32	6.3	3100	1	Sequence 3, Appli
					Sequence 1, Appli

C	28	32	6.3	10144	4	US-10-204-708-94
	29	31.8	6.2	809	3	US-09-276-531-27
	30	31.8	6.2	2207	3	US-08-956-322-3
	31	31.8	6.2	2329	3	US-08-956-322-1
	32	31.8	6.2	3350	3	US-09-110-116-2
	33	31.6	6.2	19233	4	US-10-204-708-45
	34	31.4	6.2	1024	4	US-09-328-475C-20
	35	31.4	6.2	1830121	4	US-09-557-884-1
	36	31.4	6.2	1830121	4	US-09-643-990A-1
	37	31.2	6.1	3211	2	US-08-574-959A-8
	38	31.2	6.1	3211	3	US-09-357-014-8
	39	31.2	6.1	3901	2	US-08-574-959A-6
	40	31.2	6.1	3901	3	US-09-357-014-6
	41	31	6.1	678	4	US-09-328-352-713
	42	31	6.1	891	4	US-09-134-001C-139
	43	31	6.1	5163	3	US-08-700-651-1
	44	31	6.1	5163	3	US-08-928-361B-4
	45	31	6.1	5163	4	US-09-588-995A-4

#### ALIGNMENTS

#### RESULT 1

US-08-476-120-7  
Sequence 7, Application US/08476120  
Patent No. 6025143  
GENERAL INFORMATION:  
APPLICANT: Rosinski-Chupin, Isabelle R.  
APPLICANT: Tronik, Diana  
APPLICANT: Rougeon, Francois  
APPLICANT: Seidah, Nabil  
TITLE OF INVENTION: Peptides and Polypeptides Derived  
from the Submaxillary Gland of the Rat, Corresponding  
TITLE OF INVENTION: Antibodies, Corresponding Hybridomas, and Uses of  
TITLE OF INVENTION: These Products for Diagnosis, Detection, or  
TITLE OF INVENTION: Therapeutic Purposes  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/476,120  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/153,277  
FILING DATE: 17-NOV-1993  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/499,276  
FILING DATE: 19-JUL-1990  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR88/00523  
FILING DATE: 11-OCT-1989  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 88/13353  
FILING DATE: 11-OCT-1988  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300

Sequence 94, Appli  
Sequence 27, Appli  
Sequence 3, Appli  
Sequence 1, Appli  
Sequence 2, Appli  
Sequence 45, Appli  
Sequence 20, Appli  
Sequence 1, Appli  
Sequence 8, Appli  
Sequence 6, Appli  
Sequence 6, Appli  
Sequence 713, App  
Sequence 139, App  
Sequence 1, Appli  
Sequence 4, Appli  
Sequence 4, Appli

REFERENCE/DOCKET NUMBER: 004900-129  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 658 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Rat  
TISSUE TYPE: Submaxillary Gland  
CELL TYPE: Glandular  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: SMR1  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 73...510  
US-08-476-120-7

Query Match 100.0%; Score 510; DB 3; Length 658;  
Best Local Similarity 100.0%; Pred. No. 3.4e-146;  
Matches 510; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAACGTACGACGACGAGCTTCTGACGAGCATTTCCCGCTCAGAGATTCTCCAG 60  
DB 1 AAACGTACGACGACGAGCTTCTGACGAGCATTTCCCGCTCAGAGATTCTCCAG 60

QY 61 GGGCTACCAAGATGAAGTCACTGATTTGATTTGGCTGTGATCTCTTAGCATGC 120  
DB 61 GGGCTACCAAGATGAAGTCACTGATTTGATTTGGCTGTGATCTCTTAGCATGC 120

QY 121 TTCCAGTCAGGTGAGGGTGTGAGAGCCCAAGAGACAAATATCTAGAGACAA 180  
DB 121 TTCCAGTCAGGTGAGGGTGTGAGAGCCCAAGAGACAAATATCTAGAGACAA 180

QY 181 GATCTTCAACTTCTCCCTCATTTATCTTGGTCTTCAGCCTGATCCCAATGGTGACAAATA 240  
DB 181 GATCTTCAACTTCTCCCTCATTTATCTTGGTCTTCAGCCTGATCCCAATGGTGACAAATA 240

QY 241 GGAGTAACAATCACTATACCTTAATCTTCAACCACTCGTGTCTTCTTAAATCTTCCC 300  
DB 241 GGAGTAACAATCACTATACCTTAATCTTCAACCACTCGTGTCTTCTTAAATCTTCCC 300

QY 301 GGTTTTATCACTGGACCACTGTTGTGTACAGGTACCACTGATCAATATCAATCACTGG 360  
DB 301 GGTTTTATCACTGGACCACTGTTGTGTACAGGTACCACTGATCAATATCAATCACTGG 360

QY 361 CAGCTAACTGCTCCAGACCTTACCTCTAAGCAATCTCTCACTCACTTCACTTCCACA 420  
DB 361 CAGCTAACTGCTCCAGACCTTACCTCTAAGCAATCTCTCACTCACTTCACTTCCACA 420

QY 421 GAACAAGCAATACAAAAACAGATGCGCAAAATCTCCAAACACTCTCGACTACCCAAAAT 480  
DB 421 GAACAAGCAATACAAAAACAGATGCGCAAAATCTCCAAACACTCTCGACTACCCAAAAT 480

QY 481 TCCACTGATATTTTGAAGTGGTGCAAA 510  
DB 481 TCCACTGATATTTTGAAGTGGTGCAAA 510

RESULT 2  
US-09-636-215-752/c  
; Sequence 752, Application US/09636215  
; Patent No. 6620922  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Harlocker, Susan L.  
; APPLICANT: Jiang, Yuqui

APPLICANT: Henderson, Robert A.  
APPLICANT: Kalos, Michael D.  
APPLICANT: Fanger, Gary R.  
APPLICANT: Retter, Marc W.  
APPLICANT: Stolk, John A.  
APPLICANT: Day, Craig H.  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Carter, Darrick  
APPLICANT: Li, Samuel  
APPLICANT: Wang, Aijun  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Hepler, William  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
FILE REFERENCE: 210121.42717C17  
CURRENT APPLICATION NUMBER: US/09/636,215  
NUMBER OF SEQ ID NOS: 852  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 752  
LENGTH: 644  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(644)  
OTHER INFORMATION: n=A,T,C or G  
US-09-636-215-752

Query Match 7.1%; Score 36; DB 4; Length 644;  
Best Local Similarity 58.3%; Pred. No. 0.23;  
Matches 63; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 384 ACCTTACGCAATCTCTTACTCACTTCACTTCCACGACGACGAAATACAAACAGA 443  
DB 126 ACCTCAAAGGCAAACTCAACTCAAAGGAAACCCGACGACGACGACCAAC 67

QY 444 TCCCAAAATCTCCAACTACTTGGCACTACCCAAAATTCCTGATAT 491  
DB 66 TCGCAAGCCTCATACCAGGTTCTTACGACCCCAAAATCCACCAAT 19

RESULT 3  
US-09-685-166A-752/c  
; Sequence 752, Application US/09685166A  
; Patent No. 6630305  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Harlocker, Susan L.  
; APPLICANT: Jiang, Yuqui  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Stolk, John A.  
; APPLICANT: Day, Craig H.  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Carter, Darrick  
; APPLICANT: Li, Samuel  
; APPLICANT: Wang, Aijun  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Hepler, William  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
FILE REFERENCE: 210121.427C21  
CURRENT APPLICATION NUMBER: US/09/685,166A  
CURRENT FILING DATE: 2000-10-10  
NUMBER OF SEQ ID NOS: 898  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 752  
LENGTH: 644

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(644)
; OTHER INFORMATION: n=A,T,C or G
US-09-685-166A-752

Query Match
Best Local Similarity 7.1%; Score 36; DB 4; Length 644;
Matches 63; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 384 ACCTCTAAGCAATCCTCTACTCACTTCACTTCCACAGAACAGCAATACAAACAGA 443
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
126 ACCTCAAGGCAAACTCACTCAAAAGGAACCCGACGAGGACACCAACAAC 67
QY 444 TGCAGAAATCTCCAACTACTCGGACTACCCAAAATTCACCTGATAT 491
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
66 TCGCAAGACCTCATACCAGGTTCCTACGACCCCAAAATCCACCCAAT 19

RESULT 4
US-09-621-976-2813
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S. J.Y.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

Query Match
Best Local Similarity 7.0%; Score 35.8; DB 4; Length 832;
Matches 31; Conservative 115; Mismatches 107; Indels 0; Gaps 0;

QY 186 TTCACTCTCTCTCATATCTTGGTCTTCAGCTGATCCCAATGGTGGACAAATAGAGT 245
Db ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
49 WKYKTTWYRYAMWGYKKAMCTRTKXKKKGYWYWGWRYSYMAWTRWTG 108
QY 246 AACAATCACTATACCTTAAATCTTCAACACCTCGTGTCTTGTAAATCTTCCCGGTTT 305
Db || : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
109 YAYRSMYMWRYRCWKKAYRKTTCYSSKGTWWRKWKAWTTWWKKTYWAATRYW 168
QY 306 TATCACTGACCAACCATTCGTTGTTCAAGGTACCACTGAATATCAATATCAGTGGCAGT 365
Db || : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
169 WMCWTKRWASWYCWYWKARKWTKRKSRYASARSAKRCYCSWGSWKYTMWRM 228
QY 366 AACTGCTCAGACCCCTACACCTTAAGCAATCCTCTACTCACTTCAATCCACAGACA 425
Db || : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
229 WRWRGWATGAGMKAWRASCMWRKRYAGKSKTSYKSWMCWTRSKYCYTKARTGYCYR 288
QY 426 AGCAATACAAA 438
Db || : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
289 KGGMWKGRGWWYA 301

RESULT 5
US-09-328-352-2584
; Sequence 2584, Application US/09328352
; Patent No. 6562958
```

```
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2584
; LENGTH: 864
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-2584

Query Match
Best Local Similarity 6.9%; Score 35; DB 4; Length 864;
Matches 98; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 104 GGATCCTTTCAGCATGCTTCCAGTCAGGTGAGGTGTTCAGAGGCCCAAGAACACATA 163
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
397 GAATTTTATATGAAGCTTTTACTGAAATGAATGAGACATTATTGTATCGCAAAATTA 456
QY 164 ATCCTAGAGACACACAGATCCTTCAACTTTCCTCATTTCTTCTTCTTCTTCTTCTTCT 223
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
457 ATTTTGAAGAGAGAGAGTCTTCTTCCATTCCTGATTCATTTGTTATTACCTCATTTT 516
QY 224 CCAATGGTGACAAATAGGAGTAACAATCACTATACCTTAAATCTTCAACCACTCGTG 283
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
517 ATTTAGAGAAATACAAAGCCTCAATATTATATAAGACTTTTAAACCTGATCGAT 576
QY 284 TTCTTGTAAATCTTCCCGGTTTT 306
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
577 GTGATTTTCATCTCTCAGAGGGTTT 599

RESULT 6
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
```

876	GAGGCTAAACGAACAAGAGAGATTTAAAGTTCAGAAATGGTTACAGAAGTATTAAGACAGCT	935
75	GAAGTCACTGTATTTGATCTTTGGCTGTGGATCCCTTCTAGCATGCTCCAGTCAGGTGA	134
936	GTGTGGGTGTTTTTTGATTTTTTGGTTTTCTGGTTTCAATCTCGTCAITTCACAAAGATGG	995
135	GGGTGTACAGGCCCAAGAAGACCAACATAATCTCTAGAGACAACAAGATCCTTCCAACCTCT	194
996	GAGTTTTATAGAACTTAAAGCACCATGTAAGCTACTTAAAAACAACAACAAAAAGGTCA	1055
195	TGCTCAATATCTGGTCTTCAGCGCTGATCCCAATG	229
1056	TCATTTCTCAGTCTGAATTCACAAAATGCCAATG	1090

```

RESULT 8
US-10-012-542-69
; Sequence 69, Application US/10012542
; Patent No. 6627741
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: PZ029P1
; CURRENT APPLICATION NUMBER: US/10/012,542
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: EARLIER FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: 60/089,507
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089,508
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089,509
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089,510
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/090,112
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090,113
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 69
; LENGTH: 1150
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-012-542-69

```

RESULT 9  
US-09-818-512-3/c  
; Sequence 3, Application US/09818512  
; Patent No. 6537780

```

;
; TELE# 899149
;
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ9pt-Fls
;
; US-08-232-463-14
;
;
; Query Match 6.9%; Score 35; DB 1; Length 7218;
; Best Local Similarity 6.0%; Pred. No. 1.4;
; Matches 17; Conservative 148; Mismatches 118; Indels 0; Gaps 0;
;
;
; QY 183 TCCTTCAACTCTTCCTCATTTATCTTGCTCTCAGCCTGATCCAAATGGTGGCAAAATAGG 242
; 1182 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1241
;
; QY 243 AGTAACAATCACTATACCTTAATCTTCAACACCTCGTGTCTGTGTTAACTTCCCGG 302
; 1242 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1301
;
; QY 303 TTTTATCACTGGACCACTTGGTGTACAAGGTACCACTGAATATCAATATCAGTGGCA 362
; 1302 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1361
;
; QY 363 GCTAACTGCTCCAGACGCTACACCTCTAAGCAATCCTCTACTCACTTCACTTCCACAGA 422
; 1362 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1421
;
; QY 423 ACAAGCAAAATACAAAACAGATGCCAAAATCTCCAACTACT 465
; 1422 YYYYYYYYYYGGTACCAAAATCTTCTATCTCTTTAACTACT 1464
;
;

```

RESULT 7  
 US-09-461-325-69  
 ; Sequence 69, Application US/09461325A  
 ; Patent No. 6475753  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ruben et al.  
 ; TITLE OF INVENTION: 94 Human Secreted Proteins  
 ; FILE REFERENCE: P2029P1  
 ; CURRENT APPLICATION NUMBER: US/09/461,325A  
 ; CURRENT FILING DATE: 1999-12-14  
 ; EARLIER APPLICATION NUMBER: PCT/US99/13418  
 ; EARLIER FILING DATE: 1999-06-15  
 ; EARLIER APPLICATION NUMBER: 60/089,507  
 ; EARLIER FILING DATE: 1998-06-16  
 ; EARLIER APPLICATION NUMBER: 60/089,508  
 ; EARLIER FILING DATE: 1998-06-16  
 ; EARLIER APPLICATION NUMBER: 60/089,509  
 ; EARLIER FILING DATE: 1998-06-16  
 ; EARLIER APPLICATION NUMBER: 60/089,510  
 ; EARLIER FILING DATE: 1998-06-16  
 ; EARLIER APPLICATION NUMBER: 60/090,112  
 ; EARLIER FILING DATE: 1998-06-22  
 ; EARLIER APPLICATION NUMBER: 60/090,113  
 ; EARLIER FILING DATE: 1998-06-22  
 ; NUMBER OF SEQ ID NOS: 532  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 69  
 ; LENGTH: 1150  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-09-461-325-69

Query Match	6.7%	Score 34.2	DB 4	Length 1150
Best Local Similarity	47.4%	Prod. No. 1.1		
Matches 102	Conservative	0	Mismatches 113	Indels 0
Gaps	0			

```
; GENERAL INFORMATION:
; APPLICANT: BEASLEY, Ellen et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001192
; CURRENT APPLICATION NUMBER: US/09/818,512
; CURRENT FILING DATE: 2001-03-28
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 116592
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(116592)
; OTHER INFORMATION: n = A,T,C or G
US-09-818-512-3

Query Match      6.6%; Score 33.8; DB 4; Length 116592;
Best Local Similarity 52.4%; Pred. No. 12;
Matches 53; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 367 ACTGCTCCAGACCTACACCTCTAAGCAATCCCTCTACTCAACTTCATTCCACAGAACAA 426
Db 1617 ACTGCACTCAGCTAGGCAACAGCAAGACTCTGTCTCAAAAACAAACAAACAAACAA 1558

QY 427 GCATATCAAAAACAGATGCCAAAA 451
Db 1557 AACAAAACAAAACAAATAGTAAA 1533

RESULT 10
US-08-781-891-70/c
; Sequence 70, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090620tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 240052.419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5208 base pairs
; TYPE: nucleic acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-781-891-70

Query Match      6.6%; Score 33.6; DB 3; Length 5208;
Best Local Similarity 59.4%; Pred. No. 3.3;
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 329 TACAAGGTACCACTGAATATCAATATCAATGAGTGGCAGCTAACTGCTCCAGACCCCTACACCTC 388
Db 1623 TACATAGGACGTATCGTTTTCATTATCATTTGGAGATAAATGCTTAAGCATCTCCATTTC 1564

QY 389 TAAGCAATCTCTCTACTCAACTTCATTCACAGAAC 424
Db 1563 TAAATCTTCATCACTCTCAATTACATAGGACGTATC 1528

RESULT 11
US-09-791-211-11/c
; Sequence 11, Application US/09791211
; Patent No. 6448080
; GENERAL INFORMATION:
; APPLICANT: Donna T. Ward
; APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF WRN EXPRESSION
; FILE REFERENCE: RTS-0205
; CURRENT APPLICATION NUMBER: US/09/791,211
; CURRENT FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 90
; SEQ ID NO 11
; LENGTH: 5208
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (232)...(4530)
US-09-791-211-11

Query Match      6.6%; Score 33.6; DB 4; Length 5208;
Best Local Similarity 59.4%; Pred. No. 3.3;
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 329 TACAAGGTACCACTGAATATCAATATCAATGAGTGGCAGCTAACTGCTCCAGACCCCTACACCTC 388
Db 1623 TACATAGGACGTATCGTTTTCATTATCATTTGGAGATAAATGCTTAAGCATCTCCATTTC 1564

QY 389 TAAGCAATCTCTCTACTCAACTTCATTCACAGAAC 424
Db 1563 TAAATCTTCATCACTCTCAATTACATAGGACGTATC 1528

RESULT 12
US-09-618-166-70/c
; Sequence 70, Application US/09618166
; Patent No. 6583112
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed Intellectual Property Law Group
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/618,166
  FILING DATE: 17-Jul-2000
  CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
  NAME: Mcmasters, David D.
  REGISTRATION NUMBER: 33,963
  REFERENCE/DOCKET NUMBER: 240052.419C1
  TELEPHONE: (206) 622-4900
  TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 70:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 5208 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  SEQUENCE DESCRIPTION: SEQ ID NO: 70:
US-09-618-166-70
Query Match          6.6%; Score 33.6; DB 4; Length 5208;
Best Local Similarity 59.4%; Pred. No. 3.3;
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 329 TACAGGTACCACTGAATATCATATCAGTGGCAGCTACTCTCCAGACCTTACACCTC 388
Db 1623 TACATAGGACGTATCGTTTTCATTATCATTTGGAGATTAATGCTTAAGCATCTCCATTTC 1564
QY 389 TAAGCAATCCCTACTCACTCACTTCCACAGAAC 424
Db 1563 TAAATCTTCACTCACTCACTTCAATACATAGGACGTATC 1528
RESULT 13
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
```

```
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
```

OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (674435)..(674435)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (682442)..(682442)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (713652)..(713652)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (741684)..(741684)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (779455)..(779455)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (779676)..(779676)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (855539)..(855539)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (871619)..(871619)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1084830)..(1084830)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1096846)..(1096846)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1119881)..(1119881)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1130881)..(1130881)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1310988)..(1310988)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1313224)..(1313224)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1349473)..(1349473)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1349491)..(1349491)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1470091)..(1470091)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1569020)..(1569020)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1602912)..(1602912)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1603734)..(1603734)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1637998)..(1637998)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1664854)..(1664854)  
OTHER INFORMATION: n equals a, t, c, or g  
US-08-916-421B-1

Query Match  
Best Local Similarity 6.6%; Score 33.6; DB 4; Length 1664976;  
Matches 60; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 363 GCTAACTGCTCCAGACCCCTACCTCTAAGCAATCTCTTACTCACTCACTTCACTCCAGA 422  
Db 1364462 GTTATCAACTCCAGTTCCTACAGCAGCAATAATCCCTGCTATTGAAGTAAATCTACTT 1364521  
QY 423 ACAAGCAATACAAAACAGATGCCAAATCTTCAACACTACTG 466  
Db 1364522 CCATCTATTAAAGAGCAAGCCCAATATAATAAATCTCTG 1364565

RESULT 14  
US-09-105-542A-12  
; Sequence 12, Application US/09105542A  
; Patent No. 6323329  
; GENERAL INFORMATION:  
; APPLICANT: Bullerdiel, Jörn  
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES OF GENES OF THE  
; FILE REFERENCE: BOEHA 001C1CP  
; CURRENT APPLICATION NUMBER: US/09/105,542A  
; PRIOR FILING DATE: 1998-06-26  
; PRIOR APPLICATION NUMBER: PCT/DE96/02494  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 528  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)....(528)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-105-542A-12

Query Match  
Best Local Similarity 6.5%; Score 33.4; DB 4; Length 528;  
Matches 55; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 420 AGAACAAGCAATACAAAACAGATGCCAAATCTTCAACACTACTGCGACTACCCAAA 479  
Db 241 AGGAATGGAATACAAAATACATCTCAAAATCTGTAATAATCTGAAGGACCTCTA 300  
QY 480 TTCCACTGATATTTTGAAGTGGGCAAA 510  
Db 301 TGGCCAAATAATCTTGAAGAAGATGAAAAA 331

RESULT 15  
US-08-801-405B-8  
; Sequence 8, Application US/08801405B  
; Patent No. 6589750  
; GENERAL INFORMATION:  
; APPLICANT: ROUGEOT, Catherine  
; TITLE OF INVENTION: THERAPEUTIC USE OF THE SMRI PROTEIN, THE  
; SMRI MATURATION PRODUCTS, SPECIFICALLY THE QHNP  
; PENTAPEPTIDE AS WELL AS ITS BIOLOGICALLY ACTIVE  
; DERIVATIVES  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/801.405B

Mon Aug 2 07:39:40 2004

```

;
; FILING DATE: 20-Feb-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
;   NAME: Dadio, Susan M.
;   REGISTRATION NUMBER: 40,373
;   REFERENCE/DOCKET NUMBER: 012880-003
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703) 836-6620
;   TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 33 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     MOLECULE TYPE: DNA (genomic)
;     FEATURE:
;       NAME/KEY: CDS
;       LOCATION: 1..33
;   SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-08-801-405B-8

Query Match      6.5%; Score 33; DB 4; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.49;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      139 GTCAGAGCCCAAGAGACACATATCCTAGA 171
      |||||
Db       1 GTCAGAGCCCAAGAGACACATATCCTAGA 33

Search completed: July 30, 2004, 19:11:21
Job time : 313 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 13:32:45 ; Search time 838 Seconds  
(without alignments)  
2585.419 Million cell updates/sec

Title: US-09-386-850-7\_COPY\_1\_510

Perfect score: 510

Sequence: 1 AAACGACTGACGAGAGAGC.....TTTTTGAGGTGGTGGCAA 510

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq\_29Jan04.\*  
1: geneseqn1980s.\*  
2: geneseqn1990s.\*  
3: geneseqn2000s.\*  
4: geneseqn2001as.\*  
5: geneseqn2001bs.\*  
6: geneseqn2002s.\*  
7: geneseqn2003as.\*  
8: geneseqn2003bs.\*  
9: geneseqn2003cs.\*  
10: geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	510	100.0	658	2	AAQ04272
2	142	27.8	639	9	ADD29626
3	65	12.7	65	6	ABN30493
4	56.8	11.1	366	2	AAV89531
5	56.8	11.1	521	3	AAQ04378
6	53.4	10.5	153	6	ABK55154
7	47.2	9.3	1956	3	AAQ38547
8	43.6	8.5	1362	3	AAQ49080
9	43	8.4	610	2	AAV89871
10	41.2	8.1	5883	9	AD54310
11	41	8.0	5085	7	ACF62813
12	41	8.0	8085	9	ADB54279
13	39.4	7.7	65	6	ABN56280
14	39.4	7.7	56153	3	AA546793
15	38.2	7.5	1391	3	AA255632
16	38.2	7.5	6022	4	AA546661
17	38.2	7.5	40304	8	ADA03014
18	38.2	7.5	40304	9	ADB72752
19	38.2	7.5	40304	9	ADC85494
20	38.2	7.5	86574	6	ABK83560
21	38	7.5	2000	7	ADA71938
22	37.8	7.4	495	6	ABN71324
23	37.8	7.4	1870	7	ACC51126

24	37.8	7.4	110000	6	ABN71527_11
25	37.8	7.4	110000	6	ABN71527_12
26	37.4	7.3	19653	6	ABL33335
27	37	7.3	105325	6	ABK94407
28	36.4	7.1	5127	4	AAS45425
29	36.4	7.1	6222	6	ABL32693
30	36.2	7.1	5815	6	ABK40023
31	36.2	7.1	8776	6	ABK40068
32	36.2	7.1	110000	7	ABX16390_2
33	36	7.1	644	4	AAS64085
34	36	7.1	644	5	ACA59893
35	36	7.1	644	6	ABL5456
36	36	7.1	644	7	ACC95620
37	36	7.1	644	9	ADB14202
38	35.8	7.0	604	7	ABX98613
39	35.8	7.0	7029	6	ABL32620
40	35.8	7.0	7029	6	AAD28384
41	35.6	7.0	549	8	ACC73187
42	35.6	7.0	4255	6	ABL34033
43	35.2	6.9	414	5	ABV01994
44	35.2	6.9	886	4	AAI94116
45	35	6.9	864	8	ADA31297

## ALIGNMENTS

RESULT 1  
AAQ04272  
ID AAQ04272 standard; DNA; 658 BP.  
AC AAQ04272;  
XX  
DT 21-SEP-1989 (first entry)  
XX  
DE SMR1 encoding gene.  
XX  
KW Submaxillary gland; behavioural control; ss.  
XX  
OS Rattus.  
XX  
FH Key  
FT CDS  
FT Location/Qualifiers  
73..510  
/\*tag= a  
XX  
PN WO9003981-A.  
XX  
PD 19-APR-1990.  
XX  
PF 11-OCT-1988; 88FR-00013353.  
XX  
PR 11-OCT-1988; 88FR-00013353.  
XX  
PA (INSP ) INST PASTEUR.  
XX  
PI Chupin I, Tronik D, Rougeon F, Seidah N;  
XX  
DR WPI; 1990-147823/19.  
XX  
PT P-PSDB; AAR04426.  
XX  
PT New polypeptide useful therapeutically and in diagnosis - isolated from  
PT rat submaxillary gland and derived tetra: or penta:peptide(s), antibodies  
XX and hydridomas.  
XX  
PS Disclosure; Page ?; 26pp; French.  
XX  
CC This sequence encodes the SMR1 polypeptide which is secreted from rodent  
CC esp. rat submaxillary glands and is associated with control of behaviour  
CC in such animals. Abs derived from SMR1 are useful for detecting its  
CC presence in biological tissues and fluids. See also AAR04389 and AAR04420  
CC -25  
XX  
SQ Sequence 658 BP; 216 A; 162 C; 111 G; 169 T; 0 U; 0 Other;

Continuation (12 o  
Continuation (13 o  
ABL33335 Human imm  
Abk45407 DNA encod  
Aas45425 Chemical  
ABL32693 Human imm  
Abk40023 Human che  
Abk40068 Human che  
Continuation (3 of  
Aas64085 Human pro  
Aca59893 Prostate  
AB195456 Human pro  
Acc95620 Prostate  
ADB14202 Human pro  
Abx98613 Rice leaf  
ABL32620 Human imm  
Aad28384 Human che  
Acc73187 Cat flea  
ABL34033 Human imm  
ABV01994 Human pro  
AAI94116 Human neu  
ADA31297 DNA encod



XX Human; mouse; rat; splice transcript; detection; RNA transcript;  
 KW splice variant; transcriptome; oligonucleotide library; ss.  
 XX Rattus norvegicus.  
 OS WO200210449-A2.  
 XX 07-FEB-2002.  
 XX 20-JUL-2001; 2001WO-IB001903.  
 XX 28-JUL-2000; 2000US-0221607P.  
 XX 02-MAY-2001; 2001US-0287724P.  
 XX (COMP-) COMPUEN INC.  
 XX Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;  
 XX WPI; 2002-257383/30.  
 XX New oligonucleotide libraries comprising oligonucleotides which  
 PT selectively hybridize to mRNAs transcribed from a transcription unit of a  
 PT genome, useful for detecting tissue-, pathology-, and developmental-  
 PT specific genes.  
 XX Example 1; SEQ ID NO 3241; 47pp; English.  
 XX The present invention describes oligonucleotide libraries for detecting  
 CC messenger RNAs that populate a (sub-)transcriptome, where the (sub-  
 CC )transcriptome comprises messenger RNAs transcribed from multiple  
 CC transcription units that populate a genome. The library comprises several  
 CC oligonucleotides, each capable of hybridising selectively to a set of  
 CC messenger RNAs transcribed from a given transcription unit of the genome,  
 CC which encodes one or more messenger RNA splice variants. The  
 CC oligonucleotide libraries are useful for detecting mRNAs from a  
 CC biological sample, in expression profiling studies, in qualitatively or  
 CC quantitatively characterising the corresponding transcriptome, and in  
 CC detecting RNA transcripts and splice variants of human or animal  
 CC transcriptomes. The libraries may also be used as specialised mini  
 CC libraries to detect transcripts of a sub-transcriptome under a particular  
 CC biological or pathological state, and so allowing the detection of tissue  
 CC - and pathology-specific genes such as those genes only expressed in  
 CC specific tissue under a specific pathological condition; to detect  
 CC developmental specific genes; and to detect RNA transcripts and splice  
 CC variants of a transcriptome of a patient suffering from a particular  
 CC disorder. ABN27253 to ABN59589 represent oligonucleotide sequences from  
 CC rats, humans and mice, which are used in the exemplification of the  
 CC present invention. N.B. The sequence data for this patent did not form  
 CC part of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 65 BP; 12 A; 18 C; 12 G; 23 T; 0 U; 0 Other;  
 Query Match 12.7%; Score 65; DB 6; Length 65;  
 Best Local Similarity 100.0%; Pred. No. 6.6e-10;  
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 277 CCTCGTGTCTGTAACTTCCCGGTTTATCACTGGACCACTGGTGTCAAGGT 336  
 Db 1 CCTCGTGTCTGTAACTTCCCGGTTTATCACTGGACCACTGGTGTCAAGGT 60  
 QY 337 ACCAC 341  
 Db 61 ACCAC 65  
 RESULT 4  
 AAV89531  
 ID AAV89531 standard; cDNA; 366 BP.  
 XX  
 AC AAV89531;  
 XX

DT 15-FEB-1999 (first entry)  
 XX DE EST clone CP314.  
 XX Human; secreted protein; expressed sequence tag; EST; haematopoiesis;  
 KW tissue growth; activin; inhibitor; chemotaxis; chemokinesis; haemostatic;  
 KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;  
 KW gene therapy; ss.  
 XX Homo sapiens.  
 XX WO9845436-A2.  
 XX 15-OCT-1998.  
 XX 10-APR-1998; 98WO-US006955.  
 XX 10-APR-1997; 97US-00838821.  
 XX (GEMY ) GENETICS INST INC.  
 XX Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;  
 PI Spaulding V, Agostino MJ;  
 XX WPI; 1999-070077/06.  
 XX New polynucleotides encoding human secreted proteins - derived from e.g.  
 PT human blood, kidney, foetal lung, placenta, testes, brain, ovary,  
 PT pituitary, retina and colon cDNA libraries.  
 XX Claim 1; Page 247; 618pp; English.  
 XX The present sequence represents a human expressed sequence tag (EST). The  
 CC polynucleotide, which is a secreted EST, and the encoded protein are  
 CC predicted to have useful biological activities which would make them  
 CC suitable for treating, preventing or ameliorating medical conditions in  
 CC humans and animals, although no supporting data is given. Suggested  
 CC activities include nutritional activity, immune stimulating or  
 CC suppressing activity, haematopoiesis regulating activity, tissue growth  
 CC activity, activin/inhibin activity, chemotactic/chemokinetic activity,  
 CC haemostatic and thrombolytic activity, receptor/ligand activity, anti-  
 CC inflammatory activity, cadherin/tumour invasion suppressor activity,  
 CC tumour inhibition activity. The polynucleotide may also be useful for  
 CC gene therapy  
 XX Sequence 366 BP; 100 A; 100 C; 85 G; 81 T; 0 U; 0 Other;  
 Query Match 11.1%; Score 56.8; DB 2; Length 366;  
 Best Local Similarity 60.3%; Pred. No. 5e-07;  
 Matches 94; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
 QY 69 AAGATGAAGTCACGTGTTTGTATCTTTGGCTGTGGATCCTTCTAGCATGCTTCAGTC 128  
 Db 80 AAGATGAATCAGTCACGTGTTTGTATCTTTGGCTGTGGATCCTTCTAGCATGCTTCAGTC 139  
 QY 129 AGGTGAGGTGTCTAGAGCCCAAGAGACACATATCTTAGAAGACACAAAGATCCTTC 188  
 Db 140 TGGTGAGAGTCAAGAGAGCCCAAGAGACACATATCTTAGAAGACACAAAGATCCTTC 199  
 QY 189 AACTCTTCTCATATCTTGTGCTTCACCTGATCC 224  
 Db 200 ACCTTTGGCCAGGATTTGTTCCACCACCTCCTCC 235  
 RESULT 5  
 AAC04378  
 ID AAC04378 standard; cDNA; 521 BP.  
 XX  
 AC AAC04378;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX Human secreted protein 5' EST, SEQ ID NO: 8453.  
 DE

XX	Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation; gene therapy; chromosome mapping; ss.	XX	WO200212280-A2.
XX	Homo sapiens.	XX	14-FEB-2002.
XX	EPI033401-A2.	XX	30-JUL-2001; 2001WO-US023826.
XX	06-SEP-2000.	XX	03-AUG-2000; 2000US-0223265P.
XX	21-FEB-2000; 2000EP-00200610.	XX	02-OCT-2000; 2000US-0237406P.
XX	26-FEB-1999; 99US-0122487P.	XX	20-MAR-2001; 2001US-0277495P.
XX	(GBST ) GENSET.	XX	03-JUL-2001; 2001US-0302702P.
XX	Dumas Milne Edwards J, Duclert A, Giordano J;	XX	(CORI-) CORIXA CORP.
XX	WPI; 2000-500381/45.	XX	Pyle RA, Xu J, Secrist H;
XX	New nucleic acid that is a 5' expressed sequence tag (5' EST) for obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures.	XX	WPI; 2002-257462/30.
XX	Claim 1; SEQ ID NO 8453; 71pp + Sequence Listing; English.	XX	Novel polynucleotide encoding colon tumor polypeptides, useful as vaccines for treating colon cancers.
XX	The present sequence is one of a large number of 5' ESTs derived from mRNAs encoding secreted proteins. No ORF has yet been conclusively identified within the present sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and chromosome mapping procedures. They are used to obtain upstream regulatory sequences and to design expression and secretion vectors	XX	Claim 1; Page 327; 425pp; English.
XX	Sequence 521 BP; 122 A; 187 C; 87 G; 123 T; 0 U; 2 Other;	XX	The invention relates to isolated polynucleotides (I) encoding colon tumour polypeptides (II). (I) is useful for stimulating an immune response in a patient and treating colon cancer in a patient.
XX	Query Match 11.1%; Score 56.8; DB 3; Length 521;	XX	Oligonucleotides derived from (I) are useful for determining the presence of cancer in a patient. (I) and (II) are useful in pharmaceutical compositions, e.g. vaccines, and other compositions for the diagnosis and treatment of colon cancer. A composition comprising a first component selected from physiologically acceptable carriers and immunostimulants, and an antigen-presenting cell expressing (II) is useful for inhibiting development of cancer in a patient. (I) is useful in the design and preparation of ribozyme molecules for inhibiting expression of tumour polypeptides and (II). ABK54531-ABK55464 represent human colon cancer cDNA sequences of the invention
XX	Best Local Similarity 60.3%; Pred. No. 5.9e-07;	XX	Sequence 153 BP; 36 A; 43 C; 39 G; 35 T; 0 U; 0 Other;
XX	Matches 94; Conservative 0; Mismatches 62; Indels 0; Gaps 0;	XX	Query Match 10.5%; Score 53.4; DB 6; Length 153;
XX	69 AAGATGAAGTCACTGTAATTTGATCTTTGGCCTGTGGATCCTTCTAGCATGCTTCCAGTC 128	XX	Best Local Similarity 63.8%; Pred. No. 3.7e-06;
XX	100 AAGATGAATCACTGACTTGGATCTTTGGCCTTTGGCTCTTGCAGCGTGTTCACACC 159	XX	Matches 81; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
XX	129 AGGTGAGGTTTCAGAGGCCCAAGAGACACATATCTAGAGACACAGATCCTTC 188	XX	69 AAGATGAAGTCACTGTAATTTGATCTTTGGCCTGTGGATCCTTCTAGCATGCTTCCAGTC 128
XX	160 TGGTGAGAGTCAAGAGGCCCAAGAGGCCATATCCACCTGGACCGCTGCTCTCTCTCA 219	XX	26 AAGATGAATCACTGACTTGGATCTTTGGCCTTTGGGCTCTTGCAGCGTGTTCACACC 85
XX	189 AACTCTTCCTATATCTTGGTCTTCAGCCTGATCC 224	XX	129 AGGTGAGGTTTCAGAGGCCCAAGAGACACATATCTAGAGACACAGATCCTTC 188
XX	220 ACCTTTTGGCCAGGATTTGTTCCACACCTCTCTCC 255	XX	86 TGGTGAGAGTCAAGAGGCCCAAGAGGCCATATCCACCTGGACCGCTGCTCTCTCTCA 145
XX	RESULT 6	XX	189 AACTCTT 195
XX	ABK55154	XX	146 ACCTTTT 152
XX	ID ABK55154 standard; cDNA; 153 BP.	XX	RESULT 7
XX	ABK55154;	XX	AAC38547/c
XX	18-JUN-2002 (first entry)	XX	ID AAC38547 standard; DNA; 1956 BP.
XX	Human colon cancer-associated cDNA, SEQ ID No 624.	XX	XX AAC38547;
XX	Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.	XX	XX 17-OCT-2000 (first entry)
XX	Homo sapiens.	XX	Arabidopsis thaliana DNA fragment SEQ ID NO: 21366.
XX		XX	Hybridisation assay; genetic mapping; gene expression control;
XX		XX	protein identification; signal transduction pathway; metabolic pathway;
XX		XX	promoter; termination sequence; ss.
XX		XX	Arabidopsis thaliana.
XX		XX	EP1033405-A2.

XX PD 06-SEP-2000.  
XX PD 25-FEB-2000; 2000EP-00301439.  
XX PD 25-FEB-1999; 99US-0121825P.  
XX PD 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128714P.  
PR 08-APR-1999; 99US-0129845P.  
PR 16-APR-1999; 99US-0130077P.  
PR 19-APR-1999; 99US-0130449P.  
PR 21-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134356P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-014287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 21-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 23-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 26-JUL-1999; 99US-0145224P.  
PR 27-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 28-JUL-1999; 99US-0145919P.  
PR 02-AUG-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 03-AUG-1999; 99US-0146389P.  
PR 04-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 06-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 30-AUG-1999; 99US-0151080P.  
PR 31-AUG-1999; 99US-0151303P.  
PR 01-SEP-1999; 99US-0151438P.  
PR 07-SEP-1999; 99US-0151930P.  
PR 10-SEP-1999; 99US-0152363P.  
PR 13-SEP-1999; 99US-0153070P.  
PR 15-SEP-1999; 99US-0153758P.  
PR 16-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.



```
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145148P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.

PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151338P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 8.5%; Score 43.6; DB 3; Length 1362;
Best Local Similarity 48.1%; Pred. No. 0.011;
Matches 124; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 157 CAACATATCTTAGAGACACAAAGATCGTTCAACTCTTCTCATATATCTTGGTCTTCAAG 216
Db 481 CAGCAAAAGCATCAAACTCAACATCTTCTTCTCATCATCATCATCTTCTTCTTCTCAT
QY 217 CCTGATCCCAATGGTGGACAAATAGAGTAACAATCACTATACCTTAAATCTTCAACA 276
Db 421 CATCGCTCAATTTTGTCTTCTTCTGTGCCACTCTTTTGTGTTTGTGTTGTATCC 362
QY 277 CCTCGTGTCTTGTTPAATCTTCCCGTTTTATCACTGGACCACCACTGGTTGTACAAGT 336
Db 361 CATACTCTTCTGCTTCACTTCTCTTAAATAATCTTCTTCTTCTTCTTCTTCTTCTTCT 302
QY 337 ACCACTGAATATCAATATCAGTGGAGCTACTGTCTCAGACCCCTACACCTCTTAAGCAAT 396
Db 301 TATCCTCTATTCTTCTTCAATGTCTCATCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 242
QY 397 CCTCCTACTCAACTTCAT 414
Db 241 CTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
```

```
RESULT 9
ID AAV89871 standard; cDNA; 610 BP.
XX AC
XX AC AAV89871;
XX DT 15-FEB-1999 (first entry)
XX DE EST clone CV915.
XX KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
XX KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
XX KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
XX KW gene therapy; ss.
XX OS Homo sapiens.
XX PN WO9845436-A2.
XX PD 15-OCT-1998.
XX PF 10-APR-1998; 98WO-US006955.
XX PR 10-APR-1997; 97US-00838821.
XX PA (GENY) GENETICS INST INC.
XX PI Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;
XX PI Spaulding V, Agostino MJ;
XX DR WPI; 1999-070077/06.
XX The present sequence represents a human expressed sequence tag (EST). The
XX polynucleotide, which is a secreted EST, and the encoded protein are
XX predicted to have useful biological activities which would make them
XX suitable for treating, preventing or ameliorating medical conditions in
XX humans and animals, although no supporting data is given. Suggested
XX activities include nutritional activity, immune stimulating or
XX suppressing activity, haematopoiesis regulating activity, tissue growth
XX activity, activin/inhibin activity, chemotactic/chemokinetic activity,
XX haemostatic and thrombolytic activity, receptor/ligand activity, anti-
XX inflammatory activity, cadherin/tumour invasion suppressor activity,
XX tumour inhibition activity. The polynucleotide may also be useful for
XX gene therapy
XX SQ Sequence 610 BP; 169 A; 186 C; 84 G; 171 T; 0 U; 0 Other;
Query Match 8.4%; Score 43; DB 2; Length 610;
Best Local Similarity 56.8%; Pred. No. 0.011;
Matches 79; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 51 TTCTCAGAGGGGCTACCAAGATGAAGTCACTGTTATTTGATCTTTGGCTGTGGATCCT 110
DB 88 TTCTACCAAGGAGCACTTTAAAGAAATGAATTAACCTTTCTCTGGGCTGTGGCTCT 147
QY 111 TCTAGCATGCTCCAGTCAGTGGGTGTCAGAGGCCCAAGAGCAACATATCTCTAG 170
DB 148 TATTTCAATGTTTACACCCAGTGAAGTCAAGATTTCTCCAGAGCCATATCTACCTGG 207
QY 171 AAGACAAACAGATCCTTCA 189
DB 208 CCAGTGCACCACTTCCA 226
RESULT 10
ID ADB54310/0
XX ADB54310 standard; DNA; 5683 BP.
XX AC ADB54310;
XX DT 04-DEC-2003 (first entry)
XX DE Pretreated genomic DNA region 234.
XX KW colon cell proliferative disorder; non methylated CpG dinucleotide;
XX KW cytosstatic; cancer; adenoma; carcinoma; cytosine methylation state; ds.
XX OS Unidentified.
XX PN WO2003072821-A2.
XX PD 04-SEP-2003.
XX PF 27-FEB-2003; 2003WO-EP002035.
XX PR 27-FEB-2002; 2002EP-00004551.
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Adorjan P, Burger M, Maier S, Nimmrich I, Becker E, Lesche R;
XX PI Rujan T, Schmitt A;
XX DR WPI; 2003-731620/69.
XX Detecting and differentiating between colon cell proliferative disorders
XX associated with a gene or its regulatory regions comprises contacting a
XX target nucleic acid in a biological sample obtained from the subject with
XX a reagent.
XX Claim 32; SEQ ID NO 366; 74pp; English.
XX The invention relates to a novel method for detecting and differentiating
XX between colon cell proliferative disorders associated with at least one
XX gene or its regulatory regions. The method comprises contacting a target
XX nucleic acid in a biological sample obtained from the subject with at
XX least one reagent or a series of reagents, where the reagent or series of
XX reagents, distinguishes between methylated and non methylated CpG
XX dinucleotides within the target nucleic acid. The molecules of the
XX invention demonstrate cytostatic activity whilst the method may useful
XX for detecting and differentiating between colon cell proliferative
XX disorders, including cancers such as colon adenoma and colon carcinoma.
XX The PNA (peptide nucleic acid)-oligomers are useful as probes for
XX determining cytosine methylation state or single nucleotide
XX polymorphisms. The current sequence is that of the pretreated genomic DNA
XX region of the invention. This sequence is not shown within the
XX specification but is taken from Wipoweb.
XX SQ Sequence 5683 BP; 1357 A; 0 C; 1529 G; 2797 T; 0 U; 0 Other;
Query Match 8.1%; Score 41.2; DB 9; Length 5683;
Best Local Similarity 56.7%; Pred. No. 0.12;
Matches 76; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 365 TAACTGCTCCAGACCCCTACACCTCTAAGCAATCCTCTACTCACTTATTCACAGAAC 424
DB 1076 TAACTCTCCCAACCCCATCCCTTTTCAAAATATCTTAAATACCCCAACCCACATAAC 1017
QY 425 AAGCAATACAAACAGATGCCAAATCTCCACACTACTCGGACTACCCAAATTCGA 484
DB 1016 ACAAATACAAACCAACCAACCCATACACCAACCACTATATAACCAACCAACTTTT 957
QY 485 CTGATATTTTGA 498
DB 956 CTAAATAATCTAA 943
RESULT 11
ACF62813/C
ID ACF62813 standard; DNA; 5085 BP.
```



XX ACF62813;  
XX 09-OCT-2003 (first entry)  
XX Colon cancer analysis related genomic DNA SEQ ID NO:62.  
XX Human;  
XX Human; colon cancer; oestrogen receptor; myoglobin; p21; p27; p16; p53;  
XX progesterone receptor; pna; CEA; cdc2; c-erbB2; methylation; CpG;  
XX characterisation; classification; diagnosis; differentiation;  
XX colon cell proliferative disorder; gene; ds.  
XX Homo sapiens.  
XX Synthetic.  
XX WO2003014388-A2.  
XX 20-FEB-2003.  
XX 09-AUG-2002; 2002WO-EP008939.  
XX 09-AUG-2001; 2001DE-01039283.  
XX (EPIG-) EPIGENOMICS AG.  
XX Distler J, Model F, Taubert H;  
XX WPI; 2003-256600/25.  
XX Determining methylation status of CpG dinucleotides using modified  
XX genomic sequences, oligonucleotides and/or PNA-oligomers, useful in the  
XX characterization, grading, staging and/or diagnosis of colon cancer.  
XX Claim 22; Page 110-111; 219pp; English.  
XX The present invention describes a method for determining the methylation  
XX status of CpG dinucleotides within the genes for oestrogen receptor, p21,  
XX p27, p16, progesterone receptor, myoglobin, pcna, cdc2, c-erbB2, p53  
XX and/or CEA, which comprises contacting the target nucleic acid with a  
XX reagent that distinguishes between methylated and non-methylated CpG  
XX dinucleotides, and determining from the methylation status of the CpG  
XX positions the presence of a colon cancer. A set of oligomers or peptide  
XX nucleic acid (PNA)-oligomers can be used as probes for determining the  
XX cytosine methylation state and/or single nucleotide polymorphisms (SNP)  
XX of a corresponding genomic DNA by analysis of a chemically pretreated  
XX genomic DNA. The pretreated genomic DNA is useful for the determination  
XX of the methylation status of a corresponding genomic DNA and/or detection  
XX of SNPs. The methods and pretreated genomic DNA are also useful for the  
XX characterisation, classification, diagnosis and differentiation of colon  
XX cell proliferative disorders. ACF62752 to ACF63278 represent sequences  
XX used in the exemplification of the present invention  
XX  
XX Sequence 5085 BP; 1192 A; 0 C; 1271 G; 2622 T; 0 U; 0 Other;  
Query Match 8.0%; Score 41; DB 7; Length 5085;  
Best Local Similarity 54.2%; Pred. No. 0.13;  
Matches 83; Conservative 0; Mismatches 70; Indels 0; Gaps 0;  
QY 311 CTGGACCCACCATTTGTTGACAAAGGTACCACTGAATATCAATATCAGTGGCAGCTAACTG 370  
Db 3345 CTCCTCCCTCTTTATCTTTTAAACAAACAACTAAAAATCACTATATAAAAAATACACATC 3286  
QY 371 CTCACAGACCTACACCTCTAAGCAATCCTCTACTCACTCACTTCCACAGAACAGCAA 430  
Db 3285 ATCTTTAACTCAATCAATCAAAATACACAAATCACTATATAAAAAATACACATC 3226  
QY 431 ATACAAAAACAGATGCCAAAATCTCCACACTA 463  
Db 3225 AATAAACAAACACAAAATAAATACTCCACACAA 3193

RESULT 12  
ADB54279/c

ID ADB54279 standard; DNA; 8085 BP.  
XX ADB54279;  
XX 04-DEC-2003 (first entry)  
XX Pretreated genomic DNA region 203.  
XX colon cell proliferative disorder; non methylated CpG dinucleotide;  
XX cytostatic; cancer; adenoma; carcinoma; cytosine methylation state; ds.  
XX Unidentified.  
XX WO2003072821-A2.  
XX 04-SEP-2003.  
XX 27-FEB-2003; 2003WO-EP002035.  
XX 27-FEB-2002; 2002EP-00004551.  
XX (EPIG-) EPIGENOMICS AG.  
XX Adorjan P, Burger M, Maier S, Nimmrich I, Becker E, Lesche R;  
XX Rajan T, Schmitt A;  
XX WPI; 2003-731620/69.  
XX Detecting and differentiating between colon cell proliferative disorders  
XX associated with a gene or its regulatory regions comprises contacting a  
XX target nucleic acid in a biological sample obtained from the subject with  
XX a reagent.  
XX Claim 32; SEQ ID NO 335; 74pp; English.  
XX The invention relates to a novel method for detecting and differentiating  
XX between colon cell proliferative disorders associated with at least one  
XX gene or its regulatory regions. The method comprises contacting a target  
XX nucleic acid in a biological sample obtained from the subject with at  
XX least one reagent or a series of reagents, where the reagent or series of  
XX reagents, distinguishes between methylated and non methylated CpG  
XX dinucleotides within the target nucleic acid. The molecules of the  
XX invention demonstrate cytostatic activity whilst the method may be useful  
XX for detecting and differentiating between colon cell proliferative  
XX disorders, including cancers such as colon adenoma and colon carcinoma.  
XX The PNA (peptide nucleic acid)-oligomers are useful as probes for  
XX determining cytosine methylation state or single nucleotide  
XX polymorphisms. The current sequence is that of the pretreated genomic DNA  
XX region of the invention. This sequence is not shown within the  
XX specification but is taken from Wipoweb.  
XX  
XX Sequence 8085 BP; 2334 A; 0 C; 1776 G; 3974 T; 0 U; 1 Other;  
Query Match 8.0%; Score 41; DB 9; Length 8085;  
Best Local Similarity 54.2%; Pred. No. 0.16;  
Matches 83; Conservative 0; Mismatches 70; Indels 0; Gaps 0;  
QY 311 CTGGACCCACCATTTGTTGACAAAGGTACCACTGAATATCAATATCAGTGGCAGCTAACTG 370  
Db 6345 CTCCTCCCTCTTTATCTTTTAAACAAACAACTAAAAATCACTATATAAAAAATACACATC 6286  
QY 371 CTCACAGACCTACACCTCTAAGCAATCCTCTACTCACTCACTTCCACAGAACAGCAA 430  
Db 6285 ATCTTTAACTCAATCAATCAAAATACACAAATCACTATATAAAAAATACACATC 6226  
QY 431 ATACAAAAACAGATGCCAAAATCTCCACACTA 463  
Db 6225 AATAAACAAACACAAAATAAATACTCCACACAA 6193

RESULT 13  
ABN56280  
ID ABN56280 standard; DNA; 65 BP.





**This Page Blank (uspiu,**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 15:55:56 ; Search time 2522 Seconds  
(without alignments)  
8764.848 Million cell updates/sec

Title: US-09-386-850-7\_COPY\_1\_510

Perfect score: 510

Sequence: 1 AACTGACTGACACAGAGC.....TTTTGAAGTGTGGCAA 510

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl :

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg.hum.\*

31: em.htg.inv.\*

32: em.htg.other.\*

33: em.htg.mus.\*

34: em.htg.pln.\*

35: em.htg.rat.\*

36: em.htg.mam.\*

37: em.htg.vrt.\*

38: em.sy.\*

39: em.htgo.hum.\*

40: em.htgo.mus.\*

41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	510	100.0	652	10	RATSMR1 M63112 Rat SMR1 pr
2	510	100.0	658	6	A07543 R.norvegicu
3	443	86.9	563	10	RNSMR1A1 X77819 R.norvegicu
4	389.8	76.4	641	10	RNSMR1A2 X77815 R.norvegicu
5	387	75.9	160624	2	AC144418 Rattus no
6	387	75.9	241010	2	AC103030 Rattus no
7	385.4	75.6	6433	10	RATSMR1A AC103030 Rattus no
8	385.4	75.6	6533	10	RNSMR1G M59467 Rattus norv
9	385.4	75.6	8100	10	RNYCSA1 X52467 R.norvegicu
10	283.4	55.6	1618	10	RNSMR1A2G X84997 R.norvegicu
11	262.2	51.4	592	10	RNVCSA3 X77817 R.norvegicu
12	185.8	36.4	676	10	MMU82378 U82378 Mus musculu
13	179.4	35.2	676	10	MMU82377 U82377 Mus musculu
14	158.4	31.1	660	10	RNPRV1 X77816 R.norvegicu
15	155.6	30.5	680	10	MMU82380 U82380 Mus musculu
16	153	30.0	813	10	MMU82379 U82379 Mus musculu
17	151.6	29.7	743	10	MMU82376 U82376 Mus musculu
18	142	27.8	639	10	MMMSG1 X71629 M.musculus
19	142	27.8	653	10	BC034553 Mus muscu
20	142	27.8	671	10	BC055857 Mus muscu
21	139.6	27.4	655	10	BC031806 Mus muscu
22	124.2	24.4	205580	2	AC101839 Mus muscu
23	122.4	24.0	631	10	BC031921 Mus muscu
24	113.8	22.3	23003	10	MMU82375 U82375 Mus musculu
25	113.8	22.3	190544	2	AC138398 Mus muscu
26	113.8	22.3	205580	2	AC101839 Mus muscu
27	111	21.8	190544	2	AC138398 Mus muscu
28	104.8	20.5	471	10	MMMSG3 X71631 M.musculus
29	89.2	17.5	84611	2	AC139235 Mus muscu
30	88.4	17.3	61021	2	AC101916 Mus muscu
31	76.8	15.1	8239	10	RNVCS X74229 R.rattus VC
32	76.8	15.1	257158	2	AC122599 Rattus no
33	70.6	13.8	555	10	BC059094 Mus muscu
34	63.4	12.4	160624	2	AC144418 Rattus no
35	57.4	11.3	61021	2	AC101916 Mus muscu
36	56.8	11.1	366	6	AX892590 Secretd
37	56.8	11.1	521	6	BD071680 Sequence
38	56.8	11.1	521	6	BD028123 Sequence
39	56.8	11.1	723	11	G28337 human STS S
40	56.8	11.1	724	9	HUMRRP D29833 Homo sapien
41	56.8	11.1	733	9	BC015327 Homo sapi
42	56.8	11.1	957	9	AKI30485 Homo sapi
43	53.4	10.5	153	6	AX381686 Sequence
44	51	10.0	84611	2	AC139235 Mus muscu
45	47.4	9.3	195311	10	AC121585 Mus muscu

# ALIGNMENTS

RESULT 1	RATSMR1	Rat SMR1 protein mRNA, complete cds.	652 bp	mRNA	linear	ROD 27-APR-1993
LOCUS	RATSMR1	M63112 J04109				
DEFINITION	Rat SMR1 protein mRNA, complete cds.					
ACCESSION	M63112					
VERSION	M63112.1	GI:206997				
KEYWORDS	SMR1 protein.					
SOURCE	Rattus norvegicus (Norway rat)					
ORGANISM	Rattus norvegicus					
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;					
	Rattus.					
REFERENCE	1 (bases 1 to 652)					
AUTHORS	Rosinski-Chupin,I., Tronik,D. and Rougeon,F.					
TITLE	High level of accumulation of a mRNA coding for a precursor-like					

JOURNAL protein in the submaxillary gland of male rats  
MEDLINE Proc. Natl. Acad. Sci. U.S.A. 85 (22), 8553-8557 (1988)  
PUBMED 89042220  
COMMENT 3186744  
Original source text: Rat (Wistar, male) submaxillary gland, cDNA  
to mRNA, clone CD 13.  
Draft entry and computer-readable sequence for [1] kindly provided  
by I Rosinski-Chupin, 14-NOV-1988.

## FEATURES

source  
1..652  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10116"  
c1..652  
/product="SMG1 mRNA"  
73..513  
/note="SMG1 protein precursor"  
/codon\_start=1  
/protein\_id="AA42153.1"  
/db\_xref="GI:206998"  
/translation="MKSILYLIPLGLMILLACFQSGVGRGPRRQHPNRRQDPSTLPHY  
LGLQDPNGGQIGVITITPLNLPQPRVLNPGFITGPPVVGTTGYQWQLTAPD  
PTPLSNPPTQLHSTEQANTKDAKISNTTATNTQNSTDIFEGGK"  
sig\_peptide 73..126  
/note="SMG1 protein signal peptide"  
mat\_peptide 127..510  
/product="SMG1 protein"

## ORIGIN

Query Match 100.0%; Score 510; DB 10; Length 652;  
Best Local Similarity 100.0%; Pred. No. 1.8e-134;  
Matches 510; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AAATGACTGACGAGAGCTTCTGACGACGACATTTCCCGCTCAGAGTTTCTCCAAG 60  
Db 1 AAATGACTGACGAGAGCTTCTGACGACGACATTTCCCGCTCAGAGTTTCTCCAAG 60  
QY 61 GGGCTACCAAGATGAAGTCACTGTTATTTGATCTTTGGCTGTGGATCTTCTAGCATGC 120  
Db 61 GGGCTACCAAGATGAAGTCACTGTTATTTGATCTTTGGCTGTGGATCTTCTAGCATGC 120  
QY 121 TTCCAGTCAGGTGAGGCTGTGAGGCGCCCAAGAGACAAATAATCCTAGAGACAA 180  
Db 121 TTCCAGTCAGGTGAGGCTGTGAGGCGCCCAAGAGACAAATAATCCTAGAGACAA 180  
QY 181 GATCCTTCAACTCTTCTCAATATCTTGGTCTTCAACCACTCGTCTTCTGTTAATCTTCCC 240  
Db 181 GATCCTTCAACTCTTCTCAATATCTTGGTCTTCAACCACTCGTCTTCTGTTAATCTTCCC 240  
QY 241 GGAGTAACATCACTATACCTTAATCTTCAACCACTCGTCTTCTGTTAATCTTCCC 300  
Db 241 GGAGTAACATCACTATACCTTAATCTTCAACCACTCGTCTTCTGTTAATCTTCCC 300  
QY 301 GGTITTTATCACTGGACCACTTGGTGTGACAGGTACCACTGAATATCAATATCAGTGG 360  
Db 301 GGTITTTATCACTGGACCACTTGGTGTGACAGGTACCACTGAATATCAATATCAGTGG 360  
QY 361 CAGCTAATCTGCTCAGACCTTCACTCTTAAGCAATCTTCTCTACTCACTTCACTCCACA 420  
Db 361 CAGCTAATCTGCTCAGACCTTCACTCTTAAGCAATCTTCTCTACTCACTTCACTCCACA 420  
QY 421 GAACAGCAATACAAACAGATGCCAAATCTCCAACTACTGGGACTACCCAAAT 480  
Db 421 GAACAGCAATACAAACAGATGCCAAATCTCCAACTACTGGGACTACCCAAAT 480  
QY 481 TCCACTGATATTTTGAAGGTGGGCAAA 510  
Db 481 TCCACTGATATTTTGAAGGTGGGCAAA 510

## RESULT 2

A07543 658 bp RNA linear PAT 27-JUL-1993  
LOCUS

## DEFINITION

R.norvegicus SMR1 mRNA.

## ACCESSION

A07543

## VERSION

A07543.1 GI:412272

## KEYWORDS

SMR1 protein.

## SOURCE

Rattus norvegicus (Norway rat)

## ORGANISM

Rattus norvegicus

## REFERENCE

1 (Bases 1 to 658)

## AUTHORS

## TITLE

## JOURNAL

## FEATURES

## source

## CDS

## ORIGIN

## Query Match

## Best Local Similarity

## Matches 510;

## Conservative

## Mismatches

## Indels

## Gaps

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

## QY

## Db

```

RESULT 3
RNSMR1A1
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

RNSMR1A1
R.norvegicus VCS-alpha mRNA for SMR1-alpha1.
X77819
GI:732921
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1
Singer,M., Courty,Y., and Rougeon,F.
Recent evolution of genes encoding the prohormone-like protein SMR1
in the rat submandibular gland
DNA Cell Biol. 14 (2), 137-144 (1995)
95169272
7865131
PUBMED
2 (bases 1 to 563)
Courty,Y., Singer,M., Rosinski-Chupin,I. and Rougeon,F.
Episodic evolution and rapid divergence of members of the rat
multigene family encoding the salivary prohormone-like protein SMR1
Mol. Biol. Evol. 13 (6), 758-766 (1996)
96276306
8754212
PUBMED
3 (bases 1 to 563)
Courty,Y., Rosinski-Chupin,I. and Rougeon,F.
A new proline-rich protein precursor expressed in the salivary
glands of the rat is encoded by a gene homologous to the gene
coding for the prohormone-like protein SMR1
Unpublished
Location/Qualifiers
1..563
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="submandibular gland RT-PCR"
/tissue_type="submandibular gland"
/dev_stage="adult"
1..563
/gene="VCS-alpha1"
46..486
/gene="VCS-alpha1"
/codon_start=1
/product="SMR1-alpha1"
/protein_id="CAA54834.1"
/db_xref="GI:732922"
/db_xref="SPTREMBL:Q63552"
/translation="MKPLVLIPLGILIGCFOSGDRGPRROHNLRRQTPSTLRHY
LGLOPDPNDVQMRVTITIPNLQPPRLVNLPGFITGPLVVGQSTEQYQQLTAPD
PRLSHPTQLHSTEQANKTKDAKISNTTANTQNSTDIFEGGK"
112..483
/gene="VCS-alpha1"
/product="SMR1-alpha1"

mat_peptide

ORIGIN
Query Match
Best Local Similarity 86.9%; Score 443; DB 10; Length 563;
Matches 458; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 28 CAGCACATTTCCCGCTCAGAAGTTTCTCCAAAGGGGCTACCAAGATGAAGTCACTGTAT 87
Dbb|||||
1 CAGCACATTTCCCGCTCAGAAGTTTCTCCAAAGGGGCTACCAAGATGAAGTCACTGTAT 60
QY 88 TTGATCTTTGGCCTGTGGATCTCTTAGCATCTTCCAGTCAAGTGAGGGTGTACAGAGC 147
Dbb|||||
61 TTGATCTTTGGCCTGTGGATCTTATAGATCTTCCAGTCAAGGAGGGTACAGAGC 120
QY 148 CCAAGAGACACATATCTTAGAGACACACAAAGATCCCTTCAACTCTTCTCCTCATTAATCTT 207
Dbb|||||
121 CCGAAGAGACACATATCTTAGAGAGCGACAACTCCTTCAACTCTTCTCCTCATTAATCTT 180

```

```

QY 208 GGCTTCAGCCTGATCCCAATGGTGGCAAAATAGGAGTAACAATCACTATACCTTTAAAT 267
Dbb|||||
181 GGCTTCAGCCTGATCCCAATGGTGGCAAAATAGGAGTAACAATCACTATACCTTTAAAT 240
QY 268 CTTCAACCACTCGTCTCTTCTTAACTCTCCCGGTTTATCACTGAGACCACTTGGTT 327
Dbb|||||
241 CTTCAACCACTCGTCTCTTCTTAACTCTCCCGGTTTATCACTGAGACCACTTGGTT 300
QY 328 GTACAGGTACCACTGAATATCAATATCACTGAGTGGCAGCTAACTGCTCCAGACCTTACACCT 387
Dbb|||||
301 GTACAGGTGTCACAGAAATCAATATCACTGAGTGGCAGCTAACTGCTCCAGACCTTACACCT 360
QY 388 CTAAGCAATCTCTTACTCACTTCACTTCCACAGAAACAAGCAAAATAAACAAGATGCC 447
Dbb|||||
361 CTAAGCAATCTCTTACTCACTTCACTTCCACAGAAACAAGCAAAATAAACAAGATGCC 420
QY 448 AAAATCTCCAACACTACTCGGACTACCCAAATTCACATGATATTTTGAAGTGGTGGC 507
Dbb|||||
421 AAAATCTCCAACACTACTCGGACTACCCAAATTCACATGATATTTTGAAGTGGTGGC 480
QY 508 AAA 510
Dbb|||||
481 AAA 483

RESULT 4
RNSMR1A2
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

RNSMR1A2
R.norvegicus VCS-alpha2 mRNA for SMR1-alpha2.
X77815
GI:1563744
SMR1 gene.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1
Courty,Y., Singer,M., Rosinski-Chupin,I. and Rougeon,F.
Episodic evolution and rapid divergence of members of the rat
multigene family encoding the salivary prohormone-like protein SMR1
Mol. Biol. Evol. 13 (6), 758-766 (1996)
96276306
8754212
PUBMED
2 (bases 1 to 641)
Courty,Y., Rosinski-Chupin,I. and Rougeon,F.
A new proline-rich protein precursor expressed in the salivary
glands of the rat is encoded by a gene homologous to the gene
coding for the prohormone-like protein SMR1
Unpublished
Location/Qualifiers
1..641
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Wistar"
/db_xref="taxon:10116"
/clone="ZAP33"
/tissue_type="submandibular gland"
/dev_stage="immature rat (18 days old)"
1..641
/gene="VCS-alpha2"
join(1..53,54..119,120..641)
/gene="VCS-alpha2"
1..53
/gene="VCS-alpha2"
/number=1
54..119
/gene="VCS-alpha2"
/number=2
join(66..119,120..413)
/gene="VCS-alpha2"
/codon_start=1
/product="SMR1-alpha2"

JOURNAL
FEATURES
source

```

```

/protein_id="CAA54830.1"
/db_xref="GI:1563745"
/db_xref="SPTREMBL:P70676"
/translation="MKSLYLIFGLWILIGFQSGEGVGRPRNRRPDLSTLRRY
FLQSDPNGVGMETTSIPLTIQPPSRILDLPGILTIITVQLGSLTIEYHCOLTGPFG
PCLSNPPTQLHSTE"
mat_peptide
132..410
/gene="VCS-alpha2"
/product="SMR1-alpha2"
120..641
/gene="VCS-alpha2"
/number=3
polyA_signal
614..619
/gene="VCS-alpha2"
ORIGIN
Query Match 76.4%; Score 389.8; DB 10; Length 641;
Best Local Similarity 87.5%; Pred. No. 3.8e-100;
Matches 440; Conservative 0; Mismatches 57; Indels 6; Gaps 1;
QY 8 CTGACCAGAGCTTCTGACGAGCATTTCCCGCTCAGAGTTTCTCCAGGGCTAC 67
Db 1 CTGACCAGAGCTTCTGACGAGCATTTCCCGCTCAGAGTTTCTCCAGGGCTAC 60
QY 68 CAAGATCAAGTCACTGATTGATCTTTGGCTGTGGATCCTTCTAGCATGCTTCCAGT 127
Db 61 CAAGATCAAGTCACTGATTGATCTTTGGCTGTGGATCCTTCTAGCATGCTTCCAGT 120
QY 128 CAGGTGAGGGGTGTACAGAGCCCAAGAGCAACATAATCTCTAGAGACAAACAAGATCCTT 187
Db 121 CAGGTGAGGGGTGTACAGAGCCCAAGAGCAACATAATCTCTAGAGAGCCACAAGATCCTT 180
QY 188 CAACTCTTCTCTATCTTGTCTTCCGCTGATCCCAAGTGGTGAGCAAAATAGAGTAA 247
Db 181 CAACTCTTCTGCTGTTTCTGCTTCTGATCTGATCCCTAATGGTGTGCAATGGAGGAAA 240
QY 248 CAATCACTATACCTTAAATCTTCAACACCTCTGCTTCTTCTGTTCTTCTTCCCGGTTTFA 307
Db 241 CAACAGATATACCTTAACTTCTCAACACCTAGCAGATTTATGATCTTCTCGGAATTT 300
QY 308 TCATCTGACCACTTGGTGTACAGGTACCACTGAATATCAATATCATGTCGGAGCTAA 367
Db 301 TAACTATATAACAGTGTGGTCACTAGGTACCACTG-----AATATCATTTGAGCTAA 354
QY 368 CTGCTCCAGACCTACACTCTTAAGCAATCTCTCACTCACTCACTCACTCACTCACTG 427
Db 355 CTGGTCCATCTCTGACCTTGAAGCAATCTCTCACTCACTCACTCACTCACTCACTG 414
QY 428 CAAATACAAAACAGATGCCAAATCTCCACACTACTCGCACTACTCCCAAAATTCACCTG 487
Db 415 CAAATACAAAACAGATGCCAAATCTCCACACTACTCGCACTACTCCCAAAATTCACCTG 474
QY 488 ATATTTTGAAGGTGTGCAAA 510
Db 475 ATATTTTGAAGGTGTGCAAA 497

RESULT 5
AC144418
LOCUS
DEFINITION
Rattus norvegicus clone CH230-366P1, WORKING DRAFT SEQUENCE, 46
unordered pieces.
AC144418
VERSION
AC144418.1 GI:29825863
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 160624)
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Aisbrooks,S., Amin,A., Anguiano,D.,

```

Anyalbechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A., Chu, J., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hughes, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseghe, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaekelemeh, O., Okwuonu, G., Olarnpungsoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Popovic, D., Primus, E., Pu, L. L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlarczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstein, G. and Gibbs, R. A.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

2 (bases 1 to 160624)

Unpublished

Worley, K. C.

Direct Submission

Submitted (14-APR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

Project Information

Center project name: KDMF

Center clone name: CH230-366P1

Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 117279 bases at least Q40

Consensus quality: 123973 bases at least Q30

Consensus quality: 129325 bases at least Q20

Estimated insert size: 127803; sum-of-contigs estimation

Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

\*\*\*\*\*

\* NOTE: Estimated insert size may differ from sequence length



(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 46 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1226: contig of 1226 bp in length  
\* 1227 1326: gap of unknown length  
\* 1327 2508: contig of 1182 bp in length  
\* 2509 2608: gap of unknown length  
\* 2609 4039: contig of 1431 bp in length  
\* 4040 4139: gap of unknown length  
\* 4140 5356: contig of 1217 bp in length  
\* 5357 5456: gap of unknown length  
\* 5457 6529: contig of 1073 bp in length  
\* 6530 6629: gap of unknown length  
\* 6630 7800: contig of 1171 bp in length  
\* 7801 9544: contig of 1644 bp in length  
\* 9545 9644: gap of unknown length  
\* 9645 10817: contig of 1173 bp in length  
\* 10818 12521: contig of 1604 bp in length  
\* 12522 12621: gap of unknown length  
\* 12622 14406: contig of 1785 bp in length  
\* 14407 14506: gap of unknown length  
\* 14507 16291: contig of 1785 bp in length  
\* 16292 16391: gap of unknown length  
\* 16392 17664: contig of 1273 bp in length  
\* 17665 17765: gap of unknown length  
\* 17765 19397: contig of 1633 bp in length  
\* 19398 19497: gap of unknown length  
\* 19498 21655: contig of 2158 bp in length  
\* 21656 22961: contig of 1206 bp in length  
\* 22962 23061: gap of unknown length  
\* 23062 24831: contig of 1770 bp in length  
\* 24832 24931: gap of unknown length  
\* 24932 26934: contig of 2003 bp in length  
\* 26935 27034: gap of unknown length  
\* 27035 29028: contig of 1994 bp in length  
\* 29029 29128: gap of unknown length  
\* 29129 31188: contig of 2060 bp in length  
\* 31189 31288: gap of unknown length  
\* 31289 32978: contig of 1690 bp in length  
\* 32979 33078: gap of unknown length  
\* 33079 34291: contig of 1213 bp in length  
\* 34292 34391: gap of unknown length  
\* 34392 36429: contig of 2038 bp in length  
\* 36430 36529: gap of unknown length  
\* 36530 38982: contig of 2453 bp in length  
\* 38983 39082: gap of unknown length  
\* 39083 40236: contig of 1154 bp in length  
\* 40237 40336: gap of unknown length  
\* 40337 42653: contig of 2317 bp in length  
\* 42654 42753: gap of unknown length  
\* 42754 44497: contig of 1744 bp in length  
\* 44498 44597: gap of unknown length  
\* 44598 47423: contig of 2826 bp in length  
\* 47424 47523: gap of unknown length  
\* 47524 48871: contig of 1348 bp in length  
\* 48872 48971: gap of unknown length  
\* 48972 51424: contig of 2453 bp in length  
\* 51425 51524: gap of unknown length  
\* 51525 55311: contig of 3787 bp in length  
\* 55312 55411: gap of unknown length  
\* 55412 59896: contig of 4485 bp in length  
\* 59897 59996: gap of unknown length  
\* 59997 63366: contig of 3370 bp in length  
\* 63367 63466: gap of unknown length

\* 63467 66715: contig of 3249 bp in length  
\* 66716 66815: gap of unknown length  
\* 66816 69695: contig of 2880 bp in length  
\* 69696 69795: gap of unknown length  
\* 69796 73512: contig of 3717 bp in length  
\* 73513 73612: gap of unknown length  
\* 73613 77665: contig of 4053 bp in length  
\* 77666 77765: gap of unknown length  
\* 77766 82977: contig of 5212 bp in length  
\* 82978 83077: gap of unknown length  
\* 83078 86589: contig of 3512 bp in length  
\* 86590 91066: gap of unknown length  
\* 91067 91166: gap of unknown length  
\* 91167 92290: contig of 1124 bp in length  
\* 92291 92390: gap of unknown length  
\* 92391 97722: contig of 5332 bp in length  
\* 97723 97822: gap of unknown length  
\* 97823 104887: contig of 7065 bp in length  
\* 104888 104987: gap of unknown length  
\* 104988 111110: contig of 6123 bp in length  
\* 111111 112100: gap of unknown length  
\* 112101 123129: contig of 11919 bp in length  
\* 123130 123229: gap of unknown length  
\* 123230 138677: contig of 15448 bp in length  
\* 138678 138777: gap of unknown length  
\* 138778 160624: contig of 21847 bp in length.

FEATURES

Location/Qualifiers  
1..160624  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-366P1"

ORIGIN

Query Match 75.9%; Score 387; DB 2; Length 160624;  
Best Local Similarity 100.0%; Pred. No. 3.2e-99;  
Matches 387; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 CAGTCAGGTGAGGGTCTCAGAGGCCCAAGACACACATAATCTTAGAGACACACAGAT 183  
DB 139022 CAGTCAGGTGAGGGTCTCAGAGGCCCAAGACACACATAATCTTAGAGACACACAGAT 139081  
QY 184 CTTCAACTCTTCTCATTATCTTGTCTCTCAGCCCTGATCCCAATGGTGGACAAATAGGA 243  
DB 139082 CTTCAACTCTTCTCATTATCTTGTCTCTCAGCCCTGATCCCAATGGTGGACAAATAGGA 139141  
QY 244 GTACAAATCACTATACCCCTTAATCTTCAACCACTCGTGTCTTGTAACTCTCCCGGT 303  
DB 139142 GTACAAATCACTATACCCCTTAATCTTCAACCACTCGTGTCTTGTAACTCTCCCGGT 139201  
QY 304 TTTATCACTGGACCACTTGGTCTCAAGGTACCACTGATATATCAATATCACTGGACG 363  
DB 139202 TTTATCACTGGACCACTTGGTCTCAAGGTACCACTGATATATCAATATCACTGGACG 139261  
QY 364 TTAATCTGTCAGCCCTTACACCTCTAAGCAATCTCTTCACTCACTTCACTTCCACAGAA 423  
DB 139262 TTAATCTGTCAGCCCTTACACCTCTAAGCAATCTCTTCACTCACTTCACTTCCACAGAA 139321  
QY 424 CAAGCAATACAAAACAGATGCCAAATCTTCAACACTACTGGACATACCAAAATTC 483  
DB 139322 CAAGCAATACAAAACAGATGCCAAATCTTCAACACTACTGGACATACCAAAATTC 139381  
QY 484 ACTGATATTTTGAAGGTGGGCAAA 510  
DB 139382 ACTGATATTTTGAAGGTGGGCAAA 139408

RESULT 6

AC103030/c

LOCUS

DEFINITION Rattus norvegicus clone CH230-203K20, WORKING DRAFT SEQUENCE, 3  
unordered pieces.

linear

HTG 13-MAY-2003

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be whole contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GIZF  
Center clone name: CH230-203K20  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 219721 bases at least Q40  
Consensus quality: 223909 bases at least Q30  
Consensus quality: 225134 bases at least Q20  
Estimated insert size: 235387; sum-of-contigs estimation  
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 237677: contig of 237677 bp in length  
\* 237678 237777: gap of unknown length  
\* 237778 239753: contig of 1976 bp in length  
\* 239754 239853: gap of unknown length  
\* 241010: contig of 1157 bp in length.  
\* 239854 Location/Qualifiers

source  
1. .241010  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-203K20"  
misc\_feature  
1. .2642  
/note="wgs end extension  
clone end:Sp6"  
1370 - 1797  
misc\_feature  
/note="clone boundary  
clone end:Sp6  
site:EcORI  
end sequence:BH364051"  
8702. .9391  
misc\_feature  
/note="clone boundary  
clone end:T7  
site:EcORI  
end sequence:BH364050"

ORIGIN  
Query Match 75.9%; Score 387; DB 2; Length 241010;  
Best Local Similarity 100.0%; Pred. No. 3,2e-99;  
Matches 387; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 124 CAGTCAGGTGAGGGTGTCTAGAGGCCCAAGAGCAACATATCTCTAGAACACACAGAT 183  
Db 17921 CAGTCAGGTGAGGGTGTCTAGAGGCCCAAGAGCAACATATCTCTAGAACACACAGAT 17862  
QY 184 CCTTCAACTCTTCTCTATTCTTGGTCTTTCAGCCTGTATCCCAATGTTGGCAATATAGGA 243

AC103030  
VERSION AC103030.5 GI:30580679  
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 241010)  
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,  
Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D.,  
Aryalabechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,  
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
Biswal,D., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,  
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,  
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,  
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,  
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
Deigado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,  
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,  
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,  
Gebregorisi,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,  
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,  
Harvey,Y., Haviak,P., Hawes,A., Henderson,N., Hernandez,J.,  
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M.,  
Hollins,B., Howells,S., Huiyk,S., Hume,J., Idlebird,D., Jackson,A.,  
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,  
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
Lorenshehua,L., Loulsegged,H., Lozado,R.J., Lu,X., Ma,J.,  
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,  
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
Mawney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
Morgan,M., Morris,K., Morris,S., Munidas,M., Murphy,M., Nair,L.,  
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
Nwankwelu,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,  
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,  
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,  
Puafo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
Reilly,B., Reilly,J., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
Sanders,W., Saverly,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,  
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,X.,  
Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,  
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
Williams,G., Willson,R., Wleczky,R., Wooden,H., Worley,K.,  
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
Weinstock,G., and Gibbs,R.A.  
Direct Submission  
Unpublished  
2 (bases 1 to 241010)  
Worley,K.C.  
Direct Submission  
Submitted (24-NOV-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 241010)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On May 13, 2003 this sequence version replaced gi:23106487.





CDS	gene="VCSA1" Join(2235..2288,6844..6996) /gene="VCSA1" /codon_start=1 /product="P2-Val" /protein_id="CAA59356.1" /db_xref="GI:693998" /db_xref="SPTREMBL:Q63671" /translation="MKSLYLIFGLWILLACFQTAGMIGVCLPTLPLNFSKIFLKVSQFV" PCHISPTVTVRVPYSCDNFLTP" Join(2235..2288,5604..5990) /gene="VCSA1" /codon_start=1 /product="SMR1-Val" /protein_id="CAA59355.1" /db_xref="GI:693997" /db_xref="SWISS-PROT:P13432" /translation="MKSLYLIFGLWILLACFQSGEGVGRPRRQHNPRROODPSTLPHY" LGLPDPNGSQIGVITIPLNLPQFVRLVNLPGFIPPLVVGQITTEYQVQWOLTAPD" PTPUSNPPTLLSTEQANTKDAKISNTTATTONSTDIFEGGK" 2289..5603 /gene="VCSA1" /number=2 /evidence=experimental 5604..6130 /gene="VCSA1" /number=3 /evidence=experimental 5890 /gene="VCSA1" /replace="a" 6131..6843 /gene="VCSA1" /number=3 /evidence=experimental 6350..6354 /gene="VCSA1" /replace="" 6374 /gene="VCSA1" /replace="a" 6607 /gene="VCSA1" /replace="c" 6657 /gene="VCSA1" /replace="c" 6670 /gene="VCSA1" /replace="c" 6844..7403 /number=4 /evidence=experimental
CDS	Query Match Best Local Similarity 99.7%; DB 10; Length 8100; Matches 386; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  124 CAGTCAGGTGAGGGTCTCAGAGGCCCAAGAACACATATAATCTTAGAAGACAAAGAT 183       5601 CAGTCAGGTGAGGGTCTCAGAGGCCCAAGAACACATATAATCTTAGAAGACAAAGAT 5660       184 CCTTCAACTCTTCCTCATATCTTGTCTTTCAGCCGTATCCCAATGGTGGACAAATAGA 243 5661 CCTTCAACTCTTCCTCATATCTTGTCTTTCAGCCGTATCCCAATGGTGGACAAATAGA 5720       244 GTAACAATCACTATACCCCTTAATCTTCAACCACCTCGTGTCTTGTGTTAATCTTCCCGGT 303 5721 GTAACAATCACTATACCCCTTAATCTTCAACCACCTCGTGTCTTGTGTTAATCTTCCCGGT 5780       304 TTATCACTGGACCAACATGGTGTGTAAGGTACCACTGAATATCAATATCACTGGGAG 363 
ORIGIN	

```

/gene="VCS-alpha3"
/codon_start=1
/product="SMRI-alpha3"
/protein_id="CAA54833.1"
/db_xref="GI:1563749"
/db_xref="SPTREMBL:P70634"
/translation="GDRGQRQHNRRQDPSTLPHLLCLQNGVQMRLTRNLPPSTI
TVDPFPFVLCVIGPFGISGAVENGLQYTEHYQCQTAPDPRSLNPNQNLHTTEQANTK
TQDLSTNLSSTDTF"

```

ORIGIN	Query Match	51.4%;	Score 262.2;	DB 10;	Length 592;
	Best Local Similarity	82.9%;	Pred. No. 1e-63;		
	Matches 324;	Conservative 0;	Mismatches 63;	Indels 4;	Gaps 2;
QY	124	CAGTCAGGTCAGGGTGT	CAGAGGCCCAAGAGACAA	CATAATCCTGAAGACACACAGAT	183
DB	123	CAGTAAGGTGAGAGTGC	CAGAGGCCGACAGAGACAC	ATAATCCTAGACACACACAGAT	182
QY	184	CCCTCAACTCTTCCCTCA	TTATCTTGGTCTTCAGCC---	TCATCCCAATGGTGACAAATA	240
DB	183	CCCTCAACTCTTCCCTCA	TCTCTGTCTTCAGCCCCAATGGT	GTGCAATAGGACCTAACCA	242
QY	241	GGAGTAACAATCACTATA	ACCCTTAAATCTTCAACAC	CTCGTGTCTTGTTAAATCTTCCC	300
DB	243	AGAAATTTACCCCTAGT	ACAAATACAGTTGATCCAC	CTCTTCTTGTTGTGTTATCTCT	302
QY	301	GGTTTTATCACTGGACCA	CCATTGGTTGTACAAAGGTAC	CACTGAATATCAATATCAGTGG	360
DB	303	GGTTTTATCAGTGGAGC	AGTATTCCAATCAACTAGGTAC	CAACTGAATATCATTATCAGTGT	362
QY	361	CAGCTAACTGTCAGACCC	TACACCTCTAAGCAATCCT	CCTACTCACTTCACTTCCACA	420
DB	363	CAGCTAACTGTCAGACCC	TAGATCTCTAAGCAATCCT	CTTAATCAACTTCATACCACA	422
QY	421	GAAACAAGCAAAATACAAA	ACAGATGCCAAAATCTTCCAA	CACTACTGGGACTACCCAAAAT	480
DB	423	GAAACAAGCAAAATACAAA	ACAGATGACAAAATCTTCCAA	CACTACTGGGACTACCCAAAAT	482
QY	481	TCCACTGATA-TTTTTGA	AGGTGGTGGCAA	510	
DB	483	TCCACTGATACTTTTTGA	AGGTGGTGGCAA	513	

```

RESULT 12
MMU82378
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
Source

MMU82378
Mus musculus MSG3delta salivary protein (Vcs2) mRNA, complete cds.
676 bp mRNA linear
ROD 31-DEC-1997
U82378
Mus musculus MSG3delta salivary protein (Vcs2) mRNA, complete cds.
676 bp mRNA linear
ROD 31-DEC-1997
U82378.1
GI:2289108
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 676)
Senarate-Pose,M. and Rougeon,F.
The mouse Vcs2 gene is a composite structure which evolved by gene
fusion and encodes five distinct salivary mRNA species
Gene 201 (1-2), 75-85 (1997)
98072432
9409774
2 (bases 1 to 676)
Senarate-Pose,M. and Rougeon,F.
Direct Submission
Submitted (16-DEC-1996) Immunologie, Institut Pasteur, 25 rue du
Dr. Roux, Paris 75015, France
Location/Qualifiers
1.. 676
/organism="Mus musculus"
/mol_type="mRNA"
/strain="BALB/c"
/db_xref="taxon:10090"
Source

```

Query Match	55.6%;	Score 283.4;	DB 10;	Length 1618;
Best Local Similarity	85.3%;	Pred. No. 9.4e-70;		
Matches 330;	Conservative 0;	Mismatches 51;	Indels 6;	Gaps 1;
Qy 124	CAGTCAGGTGAGGGTGT	CAGAGGCCCAAGAAGACAA	CATAATCTTAGAGACAA	CAAGAT 183
Db 1078	CAGTCAGGTGAGGGTGT	CAGAGGCCCAAGAAGACAA	CATAATCTTAGAGACCA	CAAGAT 1137
Qy 184	CTTTCAACTCTTCCGT	CTATATCTTTGGTCTTTCAG	CTGTATCCCAATGGTG	GACAAATAGGA 243
Db 1138	CTTTCAACTCTTCCGT	CTATATCTTTGGTCTTTCAG	CTGTATCCCAATGGTG	GACAAATAGGA 1197
Qy 244	GTAACTATCACTATAC	CCCTTAATCTTCAACCAC	CTCGTCTCTTGTTAAT	CTTCCCGGT 303
Db 1198	GAACAACCAAGTATAC	CCCTTAACTATTCAACC	CACTAGAGAAATTAT	GATCTTCTGGA 1257
Qy 304	TTTTATCACTGGACCA	CCATTGGTTGTACAAGG	TACCACTGAATATCA	TATCAGTGGCAG 363
Db 1258	ATTTTAACTATAATA	ACAGTTGGTCAACTAG	GTACCAATG-----	AATATCATTTGTCAG 1311
Qy 364	CTAACTGCTCCAGAC	CCCTACACCTCTAAGCA	ATCCTCTACTCAACT	CACTTCCACAGAA 423
Db 1312	CTAACTGGTCCACAT	CTCTGGACCTCTAAGCA	ATCCTCTACTCAACT	CACTTCCACAGAA 1371
Qy 424	CAAGCAAAATACAAAA	ACAGATGCCAAAATCT	TCCAACACTACTCG	CAGTACCCAAAATTC 483
Db 1372	TAGCAAAATACAAAA	ACAGATGCCAAAATCT	TCCAACACTACTCG	CAGTACCCAAAATTC 1431
Qy 484	ACTGATATTTTTGA	AGGTGGTGCAAA 510		
Db 1432	ACTGATATTTTTGA	AGGTGGTGCAAA 1458		

[illegible]

gene	sex="male"	tissue_type="submandibular_gland"
	1. .676	/gene="Vcs2"
CDS	88..519	/gene="Vcs2"
	/codon_start=1	/product="MSG2delta salivary protein"
	/protein_id="AAB93516.1"	/db_xref="GI:2289103"
	/translation="MPPLCLVFLGCLVLCFLSQCQRGFRGHDPTRPLSPNPSH FYQPPDNRVQISQPNIPFIEMFEQHSUNICVPPPLYLGEFEFLKPNPTHPIYILI RPDIPEPSKYIQVPRKSNATPAANNFITTATPNSTDSF"	
ORIGIN		
Query Match	36.4%;	Score 185.8; DB 10; Length 676;
Best Local Similarity	65.3%;	Pred. No. 6.5e-42;
Matches 340;	Conservative 0;	Mismatches 167; Indels 14; Gaps 4;
QY	1	AAACTGACTGACAGAGAGGCTTCTGACACGACACATTTCCCGCTCGAAGATTTCTCCAAAG 60
Db	15	AAAATAATTGAACAGACTGCTTCTGACCAGACATTCACCAGCAAGACTTCTCCAAAG 74
QY	61	GGCTAC - CAAGATGAAGTCACTGATTTTGATCTTTGGCTCTGGATCCTTCTAGCATG 119
Db	75	AGGCTACTGAAGATGAAGCCACTGTGTTTGGTCTTTGGCTCTGTGTTCTTATAGGATG 134
QY	120	CTTCCAGTCAGGTGAGGGTGTACAGAGCCCAAGAAGACAACATAATCTTAGAAGACAAC 179
Db	135	CTTCCCTTTCTAGTGAGTGCCAAAGAGGCTTCAGAGGACAGCATGATCCTCAAGACCCT 194
QY	180	AGATCCTTCAACTCTTCCTCATTTATCTTTGGTCTTTCAGCCTGATCCCAATGGTCGACAAT 239
Db	195	ATCTCCTTCAAATCCTTCTCTCATTTTATCTCAGCTGATCCCAATCGTGTACAGAT 254
QY	240	AGGAGTAACAATCACTATACCC-----TTAAATCTTCAACACCCTCGTGTCTTGT 290
Db	255	TTCTCAACCAGACAATATCCCATTTTTCATGTTTGACAAACCACATCTCTTAATATTG 314
QY	291	TAATCTTCCGGTTTATCACTGGACACCAATGGTTGTACAAGGTACCACCTGAATATCA 350
Db	315	TGTTCTCTCTCTCTTTATACCTTGGAAGAAGATTTGAAAAACTACCTCCTAATACTCA 374
QY	351	ATATCAGTGGCAGCTAACTGCTCCAGACCTCACCTCTAAGCAATCCTCCTACTCAACT 410
Db	375	TATTCTTTATATACTAATAGTCCAGATTTGAACCTCCAAGCAA---ATATATTCAACC 431
QY	411	TCATTCACAGAACAGCAAAATACAAAAACAGATGCCAAAATCTCCAAACACTACTCGGAC 470
Db	432	AGTTCCCAAGAAAAAATCAATGCAACACACAGCTGCCAATACCTTTATAACTACTGCTAC 491
QY	471	TACCCAAAATTCACCTGATA-TTTTTGAAGGTGGTGGCAAA 510
Db	492	TGCCCCCAATTCCTACTGATAGCTTTTGAAGGTGGTGGCAAA 532

RESULT 13	
MM082377	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	

```

PUBMED
2 (bases 1 to 676)
REFERENCE
AUTHORS Senorale-Pose, M. and Rougeon, F.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1996) Immunologie, Institut Pasteur, 25 rue du
Dr. Roux, Paris 75015, France
FEATURES
Location/Qualifiers
1..676
/organism="Mus musculus"
/mol_type="mRNA"
/strain="BALB/c"
/db_xref="taxon:10090"
/sex="male"
/tissue_type="submandibular gland"
1..676
/gene="Vcs2"
88..519
/gene="Vcs2"
/codon_start=1
/product="MSG2gamma salivary protein"
/protein_id="AAB9315.1"
/db_xref="GI:2289107"
/translation="MKALYMFVLWLIGCFSSSEQCRFGQHDTRPLSPNPSH
FPYQPNRQVQISQPNIIPIFMFQPHSLNICVPPPLYLGEFEFKLPNTHIPLYIL
RDIEPPSKYIQVPRKKNAPPAANFITTATAPNSTDSF"
ORIGIN

```

## ORIGIN

Query Match	35.2%; Score 179.4; DB 10; Length 676;
Best Local Similarity	64.5%; Pred.No. 4..4e-40;
Matches 336; Conservative	0; Mismatches 171; Indels 14; Gaps 4;

  

QY	1	A A A C T C A C T C A G C A G A G A G C T T C T G A C C A G C A C A T T C C C G C T C A G A A G T T T C T C C A A G	60
Db	15	A A A T A A T T G A A C A G A C T G C T C T G A C C A G C A C A T T C A C C A G C A A A G A A G C T T C T C C A A G	74
QY	61	G G S C T A C C A A A - G A T G A A G T C A C T G T A T T T G A T T C T T T G C C T C T G A T C C T T C T A G C A T G	119
Db	75	A C T C I A A A A A C G A T G A A G G C A C T G T A T A T G T A T T T G C T C T G G T T C T T A T A G A T G	134
QY	120	C T T T C A G T C A G T G A G G G T G T C A G A G C C C A A G A G A C A A C A T A A F C C T A G A G A C A A C A	179
Db	135	T T T C C T G T C T A G T G A G T G C C A A G A G G C T T C A G A G C A G C A T G A T C C T A C A G A C C A C T	194
QY	180	A G A T C C T T C A A C T C T C C C A N T A T C T T G G T C T T G A G C C T G A T C C C A A T G G T G G A C A A T	239
Db	195	A T C T C C T C C A A A T C C T T C T C A T T T T A T C C T C A G C C T G A T C C C A A T C G T G T A C A G A T	254
QY	240	A G G A G T A A C A A T C A C T A T A C C C ----- T T A A A T C T T C A A C A C C C T C G T G T T C T T C T	290
Db	255	T T C T C A A C C A G A C A A T A T C C C A T T T C A T G T T T G A C A A C C A C A T T C T T A A T A T T T G	314
QY	291	T A A T C T T C C G G T T T A T C A C T G B A C C A C A T T G T T A C A A G T P A C C A C T G A A T A T C A	350
Db	315	T G T T C T C T C T C T C T T T A T A C C T T G G A G A A A T T T G A A A A A C T A C C T C C T A A T A C T C A	374
QY	351	A T A T C A G T G G C A G C T A A C T G C T C C A G A C C C T A C A C C T C T A A G C A T C C T C C T A C T C A A C T	410
Db	375	T A T T C C T T A T A C T A T A T A G T C C A G A T A T T G A A C C T C C A A G C A A --- A T A T A T T C A A C C	431
QY	411	T C A T T C C A C A G A A C A A A A T A C A A A A C A G A T G C C A A A A T C T C C A A C A C T A C T G C G A C	470
Db	432	A G T T C C C A G A A A A A A T C A A A T G C A A C A C C A C C A G C T G C C A T A A C T T T A T A A C T A C T G C T A C	491
QY	471	T A C C C A A A A T T C C A C T G A T A - T T T T T G A A G G T G G T G G C A A A	510
Db	492	T C C C C C C A A T T C A C T A G A T G T T T T G A A G G T G G T G G C A A A	532

RESULT 14	
RNPRVB1	
LOCUS	
DEFINITION	
ACCESSION	

VERSION X77816.1 GI:471331  
KEYWORDS VCS gene.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 660)  
AUTHORS Courtney, Y., Rosinski-Chupin, I. and Rougeon, F.  
TITLE A new prolins-rich protein precursor expressed in the salivary glands of the rat is encoded by a gene homologous to the gene coding for the prohormone-like protein SMR1  
JOURNAL Unpublished  
FEATURES  
source Location/Qualifiers  
1..660  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Wistar"  
/db\_xref="taxon:10116"  
/clone="submandibular RT-PCR"  
/tissue\_type="submandibular gland"  
/dev\_stage="adult"  
1..50  
/number=1  
51..140  
/number=2  
Join(64..140,141..510)  
/gene="VCS-beta1"  
Join(64..140,141..510)  
/gene="VCS-beta1"  
/codon\_start=1  
/product="PR-Vbeta1"  
/protein\_id="CAA54831.1"  
/db\_xref="GI:471332"  
/translation="MKPLYLVFLVWLIGCFSLGSGRGRGHDPRRPPPPPPPPPP  
PPFGIGGPPGPHFGPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP  
PPANPTTRPPSPSRPPTVQANTIPAAISVTTGIAPNSDHF"  
141..660  
/number=3  
repeat\_region 199..234  
/rpt\_type=TANDEM  
repeat\_region 235..285  
/rpt\_type=TANDEM  
repeat\_region 286..324  
/rpt\_type=TANDEM  
polyA\_signal 640..645  
ORIGIN  
Query Match 31.1%; Score 158.4; DB 10; Length 660;  
Best Local Similarity 61.9%; Pred. No. 4.3e-34;  
Matches 313; Conservative 0; Mismatches 171; Indels 22; Gaps 3;  
QY 11 ACCAGAGAGCTTGTACAGAGACATTTCCCGCTGAGAGTTTCTCAAGGGGCTACC-A 69  
Db |||||  
1 ACCAGAGAGCTTGTACAGAGACATTTCCCGCTGAGAGTTTCTCAAGAGGCTACCAA 60  
QY 70 AGATGAAGTCACTGATTTGATCTTTGGCTGTGGATCTTCTAGCATGCTTCCAGTCA 129  
Db |||||  
61 AGATGAAGTCACTGATTTGATCTTTGGCTGTGGATCTTCTAGCATGCTTCTGTTCT 120  
QY 130 GGTGAGGGTGTACAGAGGCCCCAGAGACACATTAATCTTAGAGACACAAAGATCTTCA 189  
Db |||||  
121 GGTGAGTGTACAGAGGCCCCAGAGACATGATCTTCTAGAGACACACACACCTCTCT 180  
QY 190 ATCTTCTCTCAATATC-----TTGCTTTCAGCTGATCCCATGTTGGA 234  
Db |||||  
181 CT 240  
QY 235 CAATAGGAGTAAATCAATCACTATACCTTAAATCTTCAACCACTCGTGTCTTGTAAAT 294  
Db |||||  
241 CCAGGATTTGGTCCACACACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300

QY 295 CTTCCCGGTTTATCACTGAGCACCACCATGTTGTTGACAGGTACCACTGAATATCAATAT 354  
Db |||||  
301 GGTCCACACCCCTCA 360  
QY 355 CAGTGGCAGCTTAATGCTCCAGACCCTTACA-----CCTTAAGCAATCTCTCTACTCAA 408  
Db |||||  
361 CACCATAGACCTCCAGCAATCTCTTACAGAGGCTCTCTCCAGCAATCTCTCTCTCTCGA 420  
QY 409 CTTCAATCCACAGAACAGCAATACAAAACAGATGCAAAATCTCCAACTACTACTCG 468  
Db |||||  
421 CTTCTCTCCAGTACAGCAATACAAATACATACCACTGCCATATCTCTGTAATCTACTG 480  
QY 469 ACTACCCAAAATCTCACTGATATTTT 494  
Db |||||  
481 ATTGCCCAATCTCACTGATATTT 506  
RESULT 15  
MMU82380  
LOCUS Mus musculus MSG2alpha salivary protein (Vcs2) precursor, mRNA,  
DEFINITION complete cds.  
ACCESSION U82380 X71630  
VERSION U82380.1 GI:1769578  
KEYWORDS  
SOURCE  
ORGANISM Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 680)  
AUTHORS Tronik-Le Roux, D., Senorale-Pose, M. and Rougeon, F.  
TITLE Three novel SMR1-related cDNAs characterized in the submaxillary gland of mice show extensive evolutionary divergence in the protein coding region  
JOURNAL Gene 142 (2), 175-182 (1994)  
MEDLINE 94252564  
PUBMED 8194749  
REFERENCE 2 (bases 1 to 680)  
AUTHORS Senorale-Pose, M. and Rougeon, F.  
TITLE The mouse Vcs2 gene is a composite structure which evolved by gene fusion and encodes five distinct salivary mRNA species  
JOURNAL Gene 201 (1-2), 75-85 (1997)  
MEDLINE 98072432  
PUBMED 9409774  
REFERENCE 3 (bases 1 to 680)  
AUTHORS Senorale-Pose, M. and Rougeon, F.  
TITLE Direct Submission  
JOURNAL Submitted (16-DEC-1996) Immunologie, Institut Pasteur, 25 rue du Dr. Roux, Paris 75015, France  
COMMENT On Jan 9, 1997 this sequence version replaced gi:406258.  
FEATURES  
Location/Qualifiers  
1..680  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="BALB/c"  
/db\_xref="taxon:10090"  
/sex="male"  
/tissue\_type="submandibular gland"  
1..680  
/gene="Vcs2"  
88..540  
/gene="Vcs2"  
/codon\_start=1  
/product="MSG2alpha salivary protein precursor"  
/protein\_id="AAB93420.1"  
/db\_xref="GI:1769579"  
/translation="MKALYMFVLWVLIGCFSLGSGRGRGHDPRRPLAPSTHRRH  
YHLYRDPDNDHVGPPFPPLPTSSGCFKPHSLSLYAPPEIVIQSSLIIVTIPGLILT  
LSSDSPQRNCNCPTRPPRVKANIIPPTAQLDPLKILILLDSRK"  
88..132  
/sig\_peptide  
/gene="Vcs2"  
ORIGIN





**This Page Blank (uspto)**